Molecular Modelling of ABC transporters: from ATP hydrolysis to substrate transport

Ana Sofia Fernandes de Oliveira

Dissertação para a obtenção de grau de doutor em Bioquímica
pelo Instituto de Tecnologia Química e Biológica
da Universidade Nova de Lisboa

Oeiras, Agosto 2010

Esta tese teve o apoio financeiro da FCT e do FSE no âmbito do Quadro Comunitário de Apoio, BD nº SFRH/BD/21433/2005

“Life is the art of drawing sufficient conclusions from insufficient premises.”

Samuel Butler

Acknowledgements

First of all, I want to thank myself for the courage and persistence (although some people still call it stubbornness) in pursuing a scientific career against all difficulties.

A very special thanks and my sincere gratitude goes to Cláudio Soares, who has been my supervisor and friend since the beginning. He provided me with many valuable and helpful suggestions and important advice during all the years.

I would also like to thank António Baptista for the continuous support and motivation along the years and for all the stimulating discussions (not only scientific). I want to thank him for the patience spent listening to my existential scientific doubts and for the time used explaining the most complicated subjects in a way I could truly understand them.

I want to acknowledge all the members (whether past or present) of the Protein Modelling and the Molecular Simulation Laboratory: Vitor Teixeira, Nuno Micaelo, Carlos Cunha, Miguel Machuqueiro, Diana Lousa, João Damas, Sara Campos, Luis Carlos, Pedro Magalhães and Carla Baltazar for the fantastic working environment and for all the fun we have during this last years (especially at lunch time). I have no doubt that their sense of humor helped to ease the hardest moments. A very special thanks to Bruno Victor and Carla Pinheiro for their constant friendship and pertinent advices.

Also, I would like to acknowledge ITQB for the excellent working conditions and FCT for the financial support (SFRH/BD/21433/2005).

A special appreciation goes to my family who despite never fully understanding my professional choice always supported me whenever I needed.

Also, I would like to thank my children, Cuca and Gonçalinho. How can I say this without becoming too sappy? Both of you lighted my darkest moments and I am very proud of you both.

Last but never least, my biggest thanks goes to my husband, Sérgio, who supported me stoically, morally and financially during this my attempt to become a scientist. Without his help and moral encouragement, I think I have long ago given up and this thesis would not have been finished.

List of Publications

Papers presented in this thesis

1- **Oliveira, A.S.**, A.M. Baptista, and C.M. Soares, *Insights into the molecular mechanism of an ABC transporter: conformational changes in the NBD dimer of MJ0796.* J Phys Chem B, 2010. 114(16): p. 5486-96.

2- **Oliveira, A.S.**, A.M. Baptista, and C.M. Soares, *Conformational changes induced by ATP-hydrolysis in an ABC transporter: a molecular dynamics study of the Sav1866 exporter*. 2010, submitted to J Phys Chem B.

3- **Oliveira, A.S.**, A.M. Baptista, and C.M. Soares, *Inter-domain communication mechanisms in an ABC importer: A molecular dynamics study of the MalFGK$_2$E complex.* 2010, manuscript in preparation.

Papers related to this thesis

4- Damas, J.M., **Oliveira, A.S.**, A.M. Baptista, and C.M. Soares, *Structural consequences of ATP hydrolysis on ABC transporter NBDs: molecular dynamics studies on HlyB*, 2010, submitted to Proteins.

Abstract

Despite the great advances that have been made in the past decades in the ABC transporters field, the molecular mechanisms involved in transport across membranes remains largely an enigma. To date, questions regarding the molecular mechanism of transport, nucleotide hydrolysis and inorganic phosphate exit from the binding sites are still unanswered. In this thesis the dynamic behavior of several ABC transporters during the ATP-hydrolytic cycle is investigated using molecular modeling methods. The content of this thesis is compiled in three main scientific publications [1-3], corresponding to sections 3, 4 and 5, respectively. Although these three works are performed in prokaryotic family ABC transporters, it is likely that eukaryotic ones use similar mechanisms for nucleotide hydrolysis, inter-domain communication and allocrite translocation.

In the first work (section 3), we have studied the conformational changes induced by the ATP hydrolytic cycle in the isolated nucleotide binding domain (NBD) dimer of a lipid transporter (MJ0796) from *Methanococcus jannaschii.* Using Molecular Dynamics (MD) simulations, and taking as starting point the X-ray structure of the ATP-bound MJ0796 dimer, the major conformational changes induced by nucleotide hydrolysis during the transport cycle were identified with atomic detail. Several potential intermediate states of the ATP hydrolytic cycle were studied; each consisting of different combinations of nucleotide bound forms (ADP, ADP plus inorganic phosphate and ATP). From our extensive MD simulations, the zones undergoing significant rearrangements are located in very specific regions of the protein, namely in segments 11-19 and 93-124 (helical sub-domain). Finally, in the inorganic phosphate (IP) containing simulations, a possible exit way from the binding site for the IP species was observed, involving the movement of loop 38-43. Upon hydrolysis, the movement of this loop exposes the IP molecules to the solvent enhancing its probability of diffusing away from the binding site.

In the second study (section 4), we describe a molecular modeling study with the main objective of identifying the major conformational rearrangements induced by nucleotide hydrolysis in a complete model ABC exporter (Sav1866). The Sav1866 exporter is a bacterial ABC protein belonging to the family of the lipid flippase-like ABC transporters, and it is a homologue of the human multidrug resistance proteins. Comparing our MD simulations for the pre- (with ATP-bound in the binding sites) and the post-hydrolysis (ADP and IP-bound) states, we were able to show that segments undergoing significant conformational changes upon hydrolysis are not restricted to the NBDs, but extend deeply to the transmembrane external regions. In the transmembrane domains (TMDs), these segments are localized in the extracellular loops 1 and 3, the coupling helices and in the TMD helix 7, whereas in the NBDs, the conformational changes are located in the helical sub-domain region. Additionally, we report, for the first time, the atomic details associated to the NBD dimer dissociation upon hydrolysis in the context of a complete exporter. The NBD interface opening results from the withdrawal of the ABC signature motif from the nucleotide ribose ring and inorganic phosphate. Moreover, we determined, also for the first time, for the two simulated states, the adiabatic energy profiles along the transmembrane pore for one of known Sav1866 allocrites, doxorubicin. These profiles are similar between the two states and are overall downhill, from the cytoplasmatic to the extracellular side, with local energy barriers relatively small and easy to be overcome. The only difference between the two states is related to an energy barrier located in the intracellular loop (ICL) region, in the ATP-bound state, which, upon hydrolysis, becomes severely reduced facilitating allocrite passage. These adiabatic energy profiles give us new insights into the transport “powering” mechanism and allow us to understand, at a molecular level, the unidirectionality of allocrite translocation.

In the last study (section 5), we present another molecular modeling work with the objective of identifying the consequences of ATP hydrolysis and inorganic

phosphate exit in a model ABC importer, the maltose import system ($MalFGK_2E$) from *Escherichia coli*. For that purpose, we simulated the $MalFGK_2E$ importer in three intermediate states of the ATP-hydrolytic cycle: an ATP-bound, an ADP plus inorganic phosphate and an ADP-bound state. From extensive MD simulations performed for the three states, we showed that the major rearrangements during the ATP-cycle are located in the "EAA motif" region (mainly in the "coupling helices") of the TMDs and in the A-loop and in the helical sub-domain region of the NBDs. Additionally, only in the ADP-bound simulation, we were able to clarify the molecular details associated to the NBD dimer interface dissociation, which in this case, and similarly to what was observed in the Sav1866 exporter, is accompanied by the dissociation of ADP from the ABC signature motif, but not from its corresponding P-loop motif. However, the opening of one of the nucleotide binding sites did not significantly alter the transmembrane domains conformations, at least in our simulated timescale.

All the three above described works, when compared, seem to support a general inter-domain communication mechanism in ABC transporters, both for importers and exporters, in which the NBD helical sub-domain rotation and the TMD "coupling helices" rearrangements play essential key roles.

Sumário

Apesar dos enormes avanços científicos das últimas décadas na área dos transportadores ABC, até hoje, os mecanismos moleculares responsáveis pelo transporte activo através de membranas nesta família permanecem, na generalidade, um enigma. Actualmente, várias questões associadas ao mecanismo molecular de transporte, à hidrólise de nucleotidos e à saída do fosfato inorgânico aguardam uma resposta clara e conclusiva. Ao longo desta tese, o comportamento dinâmico de vários transportadores ABC irá ser estudado, com detalhe atómico, recorrendo a métodos teórico-computacionais. O conteúdo desta tese encontra-se repartido em três publicações científicas [1-3], correspondentes às secções 3, 4 e 5, respectivamente. É importante realçar que apesar dos três trabalhos reportados terem sido realizados unicamente em membros da família ABC de procariotas, é bastante provavél que os mecanismos para a hidrólise de nucleótidos, a comunicação entre domínios e a translocação de substratos nos transportadores ABC eucariotas sejam similares.

No primeiro trabalho desta tese (secção 3) reportamos as alterações conformacional induzidas durante o ciclo de hidrólise de ATP num dímero formado pelos domínios catalíticos (NBDs) de um transportador de lipídios (MJ0796). Neste estudo, utilizaram-se métodos de simulação de dinâmica molecular (MD) e foi possível determinar, com detalhe atómico, as principais alterações conformacionais induzidas pela hidrólise dos nucleótidos nos centros activos. Neste trabalho, foram estudados vários estados intermédios do ciclo catalítico de hidrólise, cada um formado por diferentes combinações de nucleótidos nos centros activos (ADP, ADP mais fosfato inorgânico ou ATP). Das simulações realizadas, pudemos observar que as zonas que sofrem maiores alterações conformacionais estão localizados em áreas muito específicas da proteína, essencialmente nos segmentos 11-19 e 93-124 (sub-domínio helicoidal). Finalmente, e com base nos sistemas simulados

que continham fosfato inorgânico (IP), foi possível identificar uma provável via de saída para os IPs resultante do movimento dos resíduos 38-43. Após hidrólise, este segmento rearranja-se e altera a sua conformação expondo os IPs ao solvente, aumentando a probabilidade de difusão destes para fora dos centros activos.

No segundo trabalho (secção 4), é descrito detalhadamente um estudo de modelação molecular num exportador ABC completo, denominado Sav1866. O principal objectivo deste trabalho é a identificação das alterações conformacionais resultantes da hidrólise de nucleótidos nos centros activos. A Sav1866 é um membro bacteriano da família ABC associado ao transporte de lípidos e é um homólogo das proteínas humanas associadas à resistência adquirida a fármacos. Após comparação do estado pré-hidrólise (com ATP ligado nos centros activos) com o estado pós-hidrólise (com ADP e IP ligado), foi possível mostrar que os zonas que sofrem maiores rearranjos após hidrolise não se restrigem aos NBDs e estendem-se às regiões mais externas dos TMDs. Nos TMDs, as regiões mais afectadas são os *loops* extracellulares 1 e 3, as hélices de acoplamento aos NBDs e a hélice transmembranar 7. Nos NBDs, as maiores alterações localizam-se no sub-domínio helicoidal. Adicionalmente, neste estudo observámos, pela primeira vez no contexto de um exportador completo, a dissociação do dímero catalítico após hidrólise. É agora claro, que a abertura da interface do dímero dos NBDs resulta do afastamento do motivo estrutural característico desta família (motivo *assinatura ABC)* do anel de ribose do nucleótido e do fosfato inorgânico. Neste trabalho também determinámos, para os dois estados simulados (pré- e pós-hidrólise), os perfis adiabáticos de energia para a doxorubicina (um dos substratos conhecidos deste exportador) ao longo do poro trnasmembranar. A análise dos perfis mostra que estes são semelhantes para os dois estados e apresentam uma descida contínua da energia, do lado intracelular para o lado extracelular. Adicionalmente, as barreiras locais de energia são relativamente baixas e

facilmente ultrapassáveis pelos alócritos. A única diferença observada entre os dois estados estudados localiza-se na região dos *loops* intracellulares (ICLs). Após hidrólise, a barreira de energia localizada nos ICLs é significativamente reduzida, facilitando assim a passagem dos alócritos. A determinação destes perfis adiabáticos de energia permitiu-nos perceber, com detalhe molecular, a unidirectionalidade do transporte de alócritos, o que ajudou a clarificar o mecânismo de transporte.

No último trabalho (secção 5), apresentamos um estudo de modelação molecular num importador ABC modelo completo ($MalFGK_2E$ de *Escherichia coli*). O objectivo deste trabalho é identificar as consequências estruturais da hidrólise de ATP e subsequente saída dos IPs. Para isso, simulamos o importador $MalFGK_2E$ em três estados intermédios do ciclo de hidrólise: um primeiro estado com ATP ligado nos centros activos, um segundo com ADP e IP e um terceiro estado com apenas ADP. Com base nas extensas simulações realizadas para os três estados, foi possível mostrar que os rearranjos conformacionais mais relevantes localizam-se no motivo estrutural “EAA” nos TMDs (especialmente nas hélices de acoplamento). Nos NBDs, as maiores alterações localizam-se no *loop* A e no sub-domínio helicoidal. Adicionalmente, nas simulações do estado ADP, conseguimos observar os detalhes moleculares associados à dissociação da interface do dímero dos NBDs. Neste caso, e de forma semelhante ao que havia sido reportado para o exportador Sav1866, a dissociação da interface dos NBDs é acompanhada pelo afastamento do ADP do motivo *assinatura ABC*, mas não do motivo estrutural *loop* P. Todavia, a abertura de um dos centros activos não alterou significativamente a conformação dos domínios transmembranares, pelo menos na escala de tempo simulada.

Os três trabalhos previamente descritos, quando comparados, parecem sugerir a existência de um mecânismo geral de comunicação entre domínios, comum a importadores e exportadores ABC. Neste mecânismo, as alterações

conformacionais do sub-domínio helicoidal (nos NBDs) e das hélices de acoplamento (nos TMDs) são essenciais.

List of abbreviations and symbols

Aminoacids

Alanine	A	Ala	Leucine	L	Leu
Arginine	R	Arg	Lysine	K	Lys
Asparagine	N	Asn	Methionine	M	Met
Aspartate	D	Asp	Phenylalanine	F	Phe
Cysteine	C	Cys	Proline	P	Pro
Glutamine	Q	Gln	Serine	S	Ser
Glutamate	E	Glu	Threonine	T	Thr
Glycine	G	Gly	Tryptophan	W	Trp
Histidine	H	His	Tyrosine	Y	Tyr
Isoleucine	I	Ile	Valine	V	Val

Abbreviations

ADP	Adenosine diphosphate
ATP	Adenosine triphosphate
CE	Continuum Electrostatic
CFTR	Cystic Fibrosis Conductance Regulator
DNA	Deoxyribonucleic acid
ESP	Electrostatic Potential Charges
FE	Free energy
FF	Force field
HlyB	Hemolysin Secretion Protein
IP	Inorganic Phosphate
k_B	Boltzmann Constant
k_b	Bond Force Constant
k_θ	Bond Angle Force Constant
k_ξ	Improper Dihedral Force Constant
k_φ	Proper Dihedral Force Constant
MBP	Maltose Binding Protein
MD	Molecular Dynamics
MG	Magnesium ion
MM	Molecular mechanics
MMC	Metropolis Monte Carlo

ns	Nanosecond
NBD	Nucleotide Binding Domain
PB	Poisson-Boltzmann
PDB	Protein Data Bank
ps	Picosecond
QM	Quantum Mechanics
RD	Regulatory Domain
RESP	Restrained ESP
RNA	Ribonucleic acid
$r_{i,j}$	Inter-atomic distance between atoms i and j
rms	Root Mean Square
SBP	Substrate Binding Protein
T	Temperature
TMD	Transmembrane Domains
TMS	Transmembrane Segments
V	Potential Energy
W_{ij}	Electrostatic interaction between site i and j

Greek Symbols

ε	Energy well depth (in Lennard–Jones Potential); dielectric constant (in electrostatics)
ε_0	Electric Permittivity of Vacuum

ε_r	Relative Permittivity of the Medium
σ	Inter-atomic distance in the Lennard-Jones potential

1. Introduction

1.1. ABC transporters

All cells and cellular compartments are separated from the external media by membranes. Then, the survival of all organisms depends on the regulation and selective passage of some specific molecules (such as ions, amino-acids and water molecules) across these lipid membranes. These transport processes, allow not only the acquisition of nutrients but also the excretion of waste products from cells. Examples of compounds commonly exchanged across membranes include metabolites (such as glucose and pyruvate), ions (such as sodium, potassium, calcium, and chloride) as well as amino acids and nucleotides [4].

A biological membrane is much more complex than a simple phospholipid bilayer and it can incorporate different lipids (including cholesterol) and membrane proteins. Membrane proteins can be classified into two broad categories: integral (or intrinsic) and peripheral (or extrinsic), based on the nature of their interactions with the membranes [4]. The integral membrane proteins are formed by one or more membrane spanning segments and, in general, contain hydrophobic residues that interact with the fatty acyl groups of the phospholipids anchoring the protein to the membrane [4]. The peripheral membrane proteins do not interact with the hydrophobic core of the phospholipid bilayer and are usually bound to the membrane indirectly, by interactions with integral membrane proteins or directly, through interactions with the lipid polar head groups [4].

Many of the membrane embedded proteins are involved in transport of ions and other molecules (that can not cross membranes by simple passive diffusion, either because of their hydrophilic nature or because of the concentration gradient). This process is normally named facilitated transport. There are two types of membrane transport: passive and active [4], and its major difference is related to whether they require an input of free energy (from ATP hydrolysis) to function.

In a simplified way, in the passive transport, the proteins simply allow the passage of the substrate molecules across membranes at a higher rate than normal diffusion. Since passive transporters do not work against a concentration gradient, no energy input is required. On the other hand, active transport requires the hydrolysis of ATP since it involves the translocation of molecules across a membrane against its concentration gradient.

The membrane transporter proteins are grouped into several families, where all members are related in sequence and have probably a common evolutionary origin and molecular mechanism. In this thesis, we will only focus in one of these families, the ABC family and we will try to answer some of the questions related with ATP-hydrolysis and with its coupling to transport.

ABC (**A**TP **B**inding **C**assette) family forms one of the largest superfamilies of proteins [5] and all members couple the energy from ATP binding and hydrolysis to a large variety of essential biological phenomena, comprising not only the pumping, against the concentration gradient, of substrates across biological membranes, [5] but also to other non-transport-related processes [6, 7]. This family exists through all phyla ranging from bacteria to man. The term ABC system was assigned because of several highly conserved sequences of residues associated with nucleotide (ATP) binding and hydrolysis [8, 9]. Although the vast majority of the ABC systems are associated with membrane transport, a few members appear to have been sequestered during evolution to serve alternative functions, such as DNA repair [10], translation elongation [11] and chromatin organization [12]. For example, one of the best known atypical ABC member is the UvrA protein, which is a cytoplasmic enzyme involved in DNA repair and is it formed by two typical ATPase domains interrupted by DNA-binding zinc fingers [13].

Generally, the ABC systems can be divided in three main functional categories [7]. The first category represents the exporters [7], which are a group of proteins that

translocates allocrites to the cell exterior. The second category, the importers [7], act as uptake systems retrieving allocrites from the cell exterior, and the third category is apparently not involved in transport but rather in alternative functions [7]. The term ABC transporters is restricted to the first and the second classes and they contain typical integral membrane domains [7].

Despite that each ABC transporter involved in substrate translocation is relatively specific for that given substrate [5], the variety of allocrites is huge, ranging from small organic and inorganic ions, mono- and oligissaccharides, amino-acids, vitamins, peptides and even to proteins, including toxins.

Until this moment, all the identified members of this family present a unidirectional allocrite transport [5] and no ABC transporter is known to be capable of working in both directions. Additionally, importers have only been found in prokaryotes while exporters are widely spread and expressed through all kingdoms of life [14].

Over the last decades, the ABC transporter family has received considerable attention, especially due to its connections to several human genetic diseases [15] and to multidrug resistance [16]. There are ~52 ABC genes identified in the human genome [17] and mutations in several of them lead to human genetic disorders, including a bleeding disorder [18], several eye [19, 20] and liver [21] diseases, and cystic fibrosis [22]. Moreover, in humans some members of this family are known by their function in the extrusion of chemotherapeutic drugs and thus contributing to cancer cells resistance [23]. Similarly, in bacteria [24] and fungi [25], ABC transporters catalyze the export of toxic substances, contributing to drug and antibiotic resistance.

Although some of the first ABC transporters [26] were identified over 30 years ago, the molecular mechanism involved in inter-domain communication, substrate recognition and transport across membranes remains largely unknown. Nowadays, there is a large amount of experimental and theoretical data available for several

ABC transporters (both importers and exporters), but, the exact atomic detailed mechanism for hydrolysis and transport is mainly an enigma.

1.2. Structure and Function of ABC transporters

Despite the large diversity of transported allocrites and the difference in the transport direction, both importers and exporters share a common minimum functional unit. A typical ABC transporter is formed by four domains [27] (see Figure 1): two hydrophobic membrane-spanning or integral membrane domains (designated as **T**rans-**M**embrane **D**omains –TMDs) and two hydrophilic domains (peripherally located at the cytoplasmic side), where nucleotide binding and hydrolysis occurs (designated as **N**ucleotide **B**inding **D**omains –NBDs).

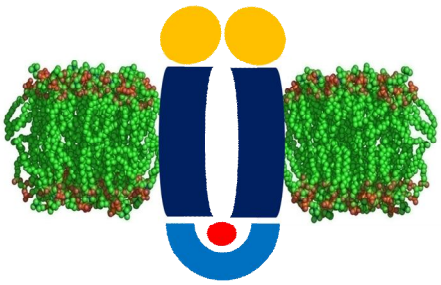


Figure 1- Minimum functional unit organization of an ABC transporter. The ABC architecture comprises two transmembrane domains (in dark blue) and two nucleotide binding domains (in yellow). Some ABC transporters receive their substrate (in red) from the bilayer whereas other ABC members receive it from the aqueous phase or from substrate binding proteins (in blue). In the last case, the substrate binding proteins prime function is the uptake and delivery of the allocrite to the transporter.

The TMD units form the transport pathway through which substrate crosses the membrane [28] and they consist of several membrane-spanning α helices (generally 8-22 helices for importers and 12 helices for exporters) [5, 28]. These domains present considerable structural variability between family members [5] and they are believed to determine the transporter substrate specificity. In contrast, the NBDs are highly conserved through the family presenting the characteristic Walker-A (or P-loop) and B motifs [5] and the ABC signature motif [5]. The Walker-A and B motifs are found in all ATPases [5] and are associated with nucleotide binding and

hydrolysis [5], whereas the ABC signature motif is a set of residues specific to the ABC family [2]

In the majority of the bacterial ABC transporters, the four basic domains are frequently coded and expressed as separated single polypeptide chains (Figure 2A) that are assembled to form a complex suitable for allocrite transport (as for example the maltose transporter $MalFGK_2$). However, the four basic domains can be fused together in almost every possible combination imagined and any domain arrangement can be found in nature (see Figure 2). For example, there are several cases in which the domains are fused together forming larger, multifunctional polypeptide chains (Figure 2B and Figure 2C). This is the case of the ribose transporter (Rbs) from *E. coli* (where the two ATP binding domains are fused together -Figure 2B) or the siderophore/haem/vitamin B_{12} transporter (where both TMDs are fused together forming only one large transmembrane polypeptide chain -Figure 2C). In the majority of the ABC exporters known, such as the lipid flippase from *E. coli* and the multidrug lipid flippase from *Staphylococcus aureaus*, the N-terminus (which comprises the hydrophobic transmembrane domain) and the C-terminus (the ATP binding cassette domain) are fused into a unique polypeptide chain (Figure 2D). Usually, these transporters are called “half-size” transporters. Additionally, in many eukaryotic ABC transporters all four domains are expressed as a unique polypeptide chain, as in the case of human multidrug resistance P-glycoprotein and CFTR (Figure 2F).

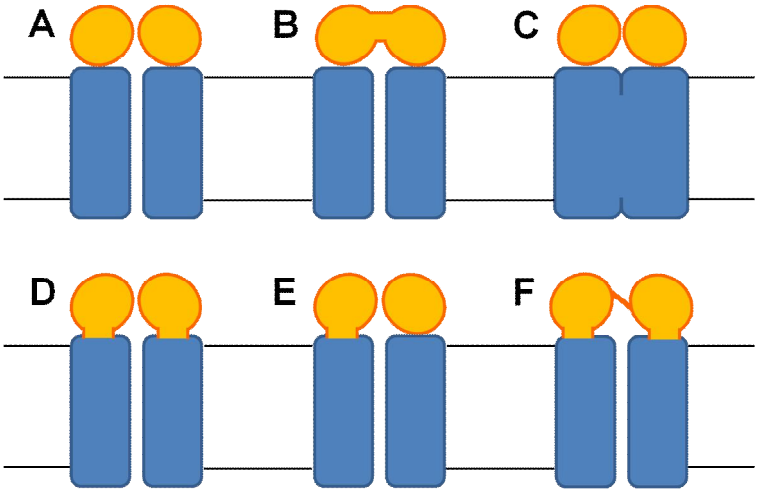


Figure 2- Possible domain architecture of ABC transporters. A typical ABC transporter is composed of two hydrophobic transmembrane domains (TMDs) and two water soluble nucleotide-binding domains (NBDs) bound to the cytosolic face of the TMDs. The TMDs form the translocation pathway and they can form a homo or a heterodimer. The NBDs are the engines of ABC transporters and they couple the ATP energy to the transport process. Schematically indicated are: TMDs (blue rectangles) and NBDs (yellow spheres). The domains are frequently expressed as separate polypeptide chains, but they can be fused in all possibilities represented in this figure. (**A**) All four basic domains are expressed as separated single polypeptide chains. An example of such archiquecture is maltose/maltodextrin transporter ($MalFGK_2$) and vitamin B_{12} transporter (BtuCD) from E. coli. (**B**) The two NBDs are fused together while the two TMDs are expressed as separated single chains. An example is the ribose transporter (Rbs) from E. coli. (**C**) The two TMDs are fused together while the two NBDs are expressed as separated polypeptide chains. An example of this archiquecture is the siderophore/haem/vitamin B_{12} transporter (Fhu). (**D**) Each TMD is fused to one NBD (frequently named half-transporter). Some examples of this architecture is the lipid flippase from E. coli (MsbA), the mitochondrial peptide transporter from Saccharomyces cerevisiae (Mdl) and the multidrug lipid flippase from Staphylococcus aureaus (Sav1866). (**E**) One TMD is fused to one NBD while the other TMD and NBD are expressed separatly as single polypeptide chains. An example is the YhiGHI transporter. (**F**) All four domains are expressed as a unique polypeptide chain. Examples of this archiquecture is the yeast pleiotropic drug resistance transporter (PDR5), the human multidrug transporter (P-glycoprotein) and the cystic fibrosis conductance regulator (CFTR).

Despite the fact that the previously described four core domains are sufficient to mediate allocrite translocation, some ABC members possess additional domains with peripheral or regulatory functions. This is the case of CFTR or the maltose transporter. CFTR presents an extra domain with regulatory functions located between the NBDs, which has no equivalence to any other known domain in the ABC family [29, 30]. MalK presents an extension in the C-terminal region of the NBDs, with regulatory and enzymatic functions (which are apparently independent of the transport process itself) [31]. Furthermore, all prokaryotic importers, in order to properly perform their function, rely on high-affinity binding proteins located in the external side of the membrane, and whose function is to retrieve the allocrites from the medium and deliver them to the transporter [32].

Despite the fact that high resolution crystal structures obtained for the isolated [33-49] and the dimeric [46, 47, 50-56] ATPase modules, together with the full-length ABC transporters structures [57-65], provided us with relevant structural information and insights into the transport cycle, there is still much to be done before we have a complete understanding of how ABC transporters work. There are still many open questions about the NBD:TMD communication mechanisms, ATP-hydrolysis or the conformational changes induced during a transport cycle in the TMDs.

In the following sub-chapters, the ABC transporters basic domains (both in importers and exporters) will be briefly described.

1.2.1. Transmembrane Domain (TMD)

The TMDs are two highly hydrophobic protein domains that form the channel through which the allocrite is translocated [28]. The TMD dimer can be either formed by two copies of the same monomer or by two distinct monomers, forming by this reason homo- or heterodimers. The TMDs are essentially formed by several α-helices that traverse the membranes. Many ABC members posses a “two-times-six” helices, in a total of 12 transmembrane segments (TMSs) per functional unit [5, 28]. However, this number may vary from 4 to 11 for each individual transporter [5, 28]. The large structural diversity of the TMDs can easily be illustrated looking at the X-ray data available for full-length transporters [57-65] (see Figure 3).

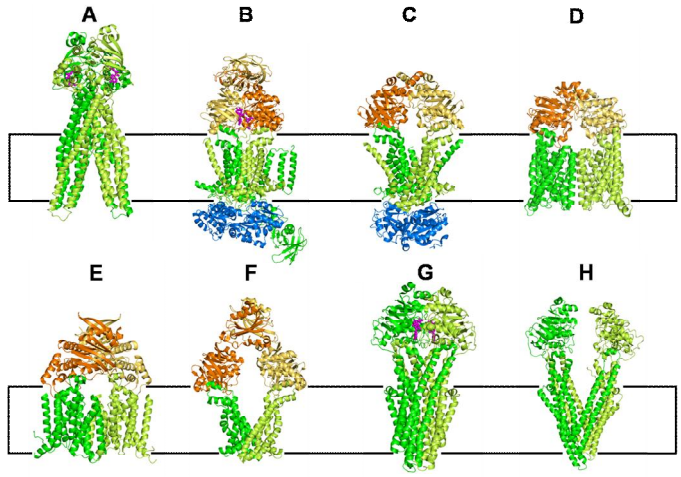


Figure 3- TMDs structural diversity. X-ray structures for **A**) Sav1866 homodimer [58]. **B**) $MalFGK_2E$ complex [62]. **C**) ModA-ModBC complex [60]. **D**) Vitamin B_{12} (BtuCD) transporter [57]. **E**) Putative metal chelate transporter H10796 [63]. **F**) MetNI methionine transporter [66]. **G**) MsbA [64]. **H**) P-glycoprotein [65]. The approximate location of the lipid bilayer is indicated by the black lines. Some structures (**B**, **C**, **F**) have an additional extension of the NBD, probably involved in regulation of transport activity. As can be seen, the transmembrane domain structures are quit variable among the ABC family members and the primary sequence is hardly conserved at all. Some of these structures were determined in the presence of ATP (or another nucleotide), some in the absence of nucleotide.

So, in this figure the nucleotide species is represented by magenta sphere whenever present. In these eight examples, only structures **B** and **H** show the transported allocrite bound to the TMDs.

TMDs present low sequence identity between family members. If we consider that all ABC transporters share a common evolutionary origin [32], the low sequence identity may imply that each TMDs is designed to satisfy different structural constraints, depending on the transported allocrite. In this case, the TMDs low sequence identity among the family most probably reflects the large diversity of transported molecules. Another factor that can help explain the large differences (whether in sequence and in structure) in the TMD region is the physiological lipid environment where each transporter is inserted.

Despite the sequence variability, there is one sequence of residues highly conserved in the ABC importers [67]. This sequence is named “EAA” motif or L-loop [67, 68], and it is usually located in one cytoplasmic loop and appropriately positioned to interact with the ATP-binding domains [69], being most probably involved in inter-domain communication [58].

1.2.2. Nucleotide Binding Domain (NBD)

The ATP-binding domains are the most characteristic feature of the ABC family and they use ATP binding and hydrolysis to power transport [5]. Presumably, ATP binding and/or hydrolysis is coupled to conformational changes in the TMDs that mediate the unidirectional pumping of allocrites across membranes [70, 71]. NBDs are highly hydrophilic domains and are always located in the cytoplasmic side of the membrane. Each NBD has more or less 200 amino-acid residues, and contrary to the TMDs, their sequence presents considerable identity, ranging from 30 to 50% [8], as can be seen in the illustrative sequence alignment shown in Figure 4.

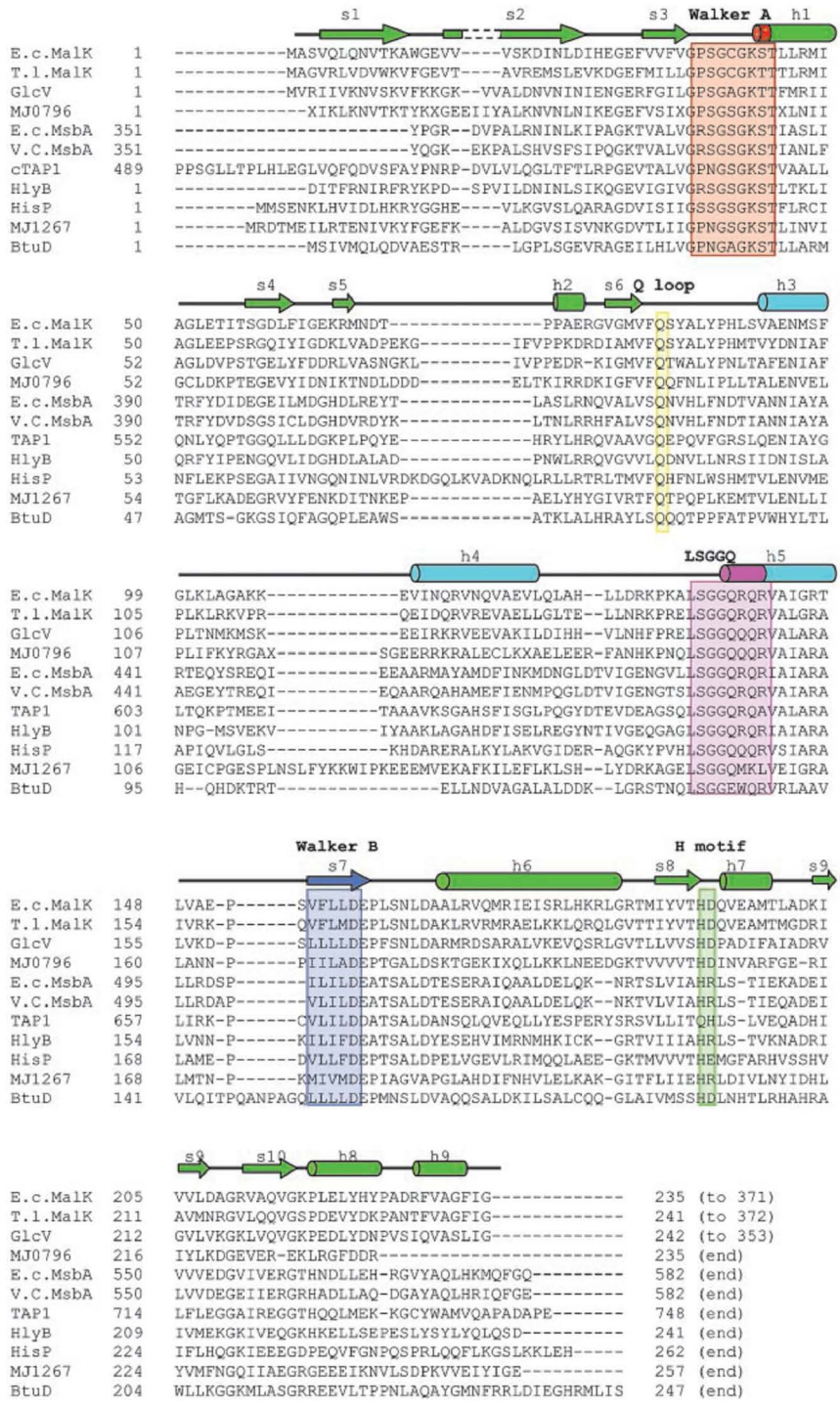

s1 s2 s3 Walker A h1
E.c.MalK 1 ----------MASVQLQNVTKAWGEVV----VSKDINLDIHEGEFVVFVGPSGCGKSTLLRMI
T.l.MalK 1 ----------MAGVRLVDVWKVFGEVT----AVREMSLEVKDGEFMILLGPSGCGKTTTLRMI
GlcV 1 ----------MVRIIVKNVSKVFKKGK--VVALDNVNINIENGERFGILGPSGAGKTTFMRII
MJ0796 1 ----------XIKLKNVTKTYKXGEEIIYALKNVNLNIKEGEFVSIXGPSGSGKSTXLNII
E.c.MsbA 351 ----------YPGR--DVPALRNINLKIPAGKTVALVGRSGSGKSTIASLI
V.C.MsbA 351 ----------YQGK--EKPALSHVSFSIPQGKTVALVGRSGSGKSTIANLF
cTAP1 489 PPSGLLTPLHLEGLVQFQDVSFAYPNRP-DVLVLQGLTFTLRPGEVTALVGPNGSGKSTVAALL
HlyB 1 ----------DITFRNIRFRYKPD--SPVILDNINLSIKQGEVIGIVGRSGSGKSTLTKLI
HisP 1 --------MMSENKLHVIDLHKRYGGHE----VLKGVSLQARAGDVISIIGSSGSGKSTFLRCI
MJ1267 1 -------MRDTMEILRTENIVKYFGEFK----ALDGVSISVNKGDVTLIIGPNGSGKSTLINVI
BtuD 1 ----------MSIVMQLQDVAESTR-----LGPLSGEVRAGEILHLVGPNGAGKSTLLARM
s4 s5 h2 s6 Q loop h3
E.c.MalK 50 AGLETITSGDLFIGEKRMNDT----------PPAERGVGMVFQSYALYPHLSVAENMSF
T.l.MalK 50 AGLEEPSRGQIYIGDKLVADPEKG---------IFVPPKDRDIAMVFQSYALYPHMTVYDNIAF
GlcV 52 AGLDVPSTGELYFDDRLVASNGKL---------IVPPEDR-KIGMVFQTWALYPNLTAFENIAF
MJ0796 52 GCLDKPTEGEVYIDNIKTNDLDDD---------ELTKIRRDKIGFVFQQFNLIPLLTALENVEL
E.c.MsbA 390 TRFYDIDEGEILMDGHDLREYT----------LASLRNQVALVSQNVHLFNDTVANNIAYA
V.C.MsbA 390 TRFYDVDSGSICLDGHDVRDYK----------LTNLRRHFALVSQNVHLFNDTIANNIAYA
TAP1 552 QNLYQPTGGQLLLDGKPLPQYE----------HRYLHRQVAAVGQEPQVFGRSLQENIAYG
HlyB 50 QRFYIPENGQVLIDGHDLALAD----------PNWLRRQVGVVLQDNVLLNRSIIDNISLA
HisP 53 NFLEKPSEGAIIVNGQNINLVRDKDGQLKVADKNQLRLLRTRLTMVFQHFNLWSHMTVLENVME
MJ1267 54 TGFLKADEGRVYFENKDITNKEP----------AELYHYGIVRTFQTPQPLKEMTVLENLLI
BtuD 47 AGMTS-GKGSIQFAGQPLEAWS----------ATKLALHRAYLSQQQTPPFATPVWHYLTL
h4 LSGGQ h5
E.c.MalK 99 GLKLAGAKK----------EVINQRVNQVAEVLQLAH--LLDRKPKALSGGQRQRVAIGRT
T.l.MalK 105 PLKLRKVPR----------QEIDQRVREVAELLGLTE--LLNRKPRELSGGQRQRVALGRA
GlcV 106 PLTNMKMSK----------EEIRKRVEEVAKILDIHH--VLNHFPRELSGGQQQRVALARA
MJ0796 107 PLIFKYRGAX----------SGEERRKRALECLKXAELEER-FANHKPNQLSGGQQQRVAIARA
E.c.MsbA 441 RTEQYSREQI----------EEAARMAYAMDFINKMDNGLDTVIGENGVLLSGGQRQRIAIARA
V.C.MsbA 441 AEGEYTREQI----------EQAARQAHAMEFIENMPQGLDTVIGENGTSLSGGQRQRVAIARA
TAP1 603 LTQKPTMEEI----------TAAAVKSGAHSFISGLPQGYDTEVDEAGSQLSGGQRQAVALARA
HlyB 101 NPG-MSVEKV----------IYAAKLAGAHDFISELREGYNTIVGEQGAGLSGGQRQRIAIARA
HisP 117 APIQVLGLS----------KHDARERALKYLAKVGIDER-AQGKYPVHLSGGQQQRVSIARA
MJ1267 106 GEICPGESPLNSLFYKKWIPKEEEMVEKAFKILEFLKLSH--LYDRKAGELSGGQMKLVEIGRA
BtuD 95 H--QHDKTRT----------ELLNDVAGALALDDK--LGRSTNQLSGGEWQRVRLAAV
Walker B H motif
s7 h6 s8 h7 s9
E.c.MalK 148 LVAE-P------SVFLLDEPLSNLDAALRVQMRIEISRLHKRLGRTMIYVTHDQVEAMTLADKI
T.l.MalK 154 IVRK-P------QVFLMDEPLSNLDAKLRVRMRAELKKLQRQLGVTTIYVTHDQVEAMTMGDRI
GlcV 155 LVKD-P------SLLLLDEPFSNLDARMRDSARALVKEVQSRLGVTLLVVSHDPADIFAIADRV
MJ0796 160 LANN-P------PIILADEPTGALDSKTGEKIXQLLKKLNEEDGKTVVVVTHDINVARFGE-RI
E.c.MsbA 495 LLRDSP-------ILILDEATSALDTESERAIQAALDELQK--NRTSLVIAHRLS-TIEKADEI
V.C.MsbA 495 LLRDAP-------VLILDEATSALDTESERAIQAALDELQK--NKTVLVIAHRLS-TIEQADEI
TAP1 657 LIRK-P------CVLILDDATSALDANSQLQVEQLLYESPERYSRSVLLITQHLS-LVEQADHI
HlyB 154 LVNN-P------KILIFDEATSALDYESEHVIMRNMHKICK--GRTVIIIAHRLS-TVKNADRI
HisP 168 LAME-P------DVLLFDEPTSALDPELVGEVLRIMQQLAEE-GKTMVVVTHEMGFARHVSSHV
MJ1267 168 LMTN-P------KMIVMDEPIAGVAPGLAHDIFNHVLELKAK-GITFLIIEHRLDIVLNYIDHL
BtuD 141 VLQITPQANPAGQLLLLDEPMNSLDVAQQSALDKILSALCQQ-GLAIVMSSHDLNHTLRHAHRA
s9 s10 h8 h9
E.c.MalK 205 VVLDAGRVAQVGKPLELYHYPADRFVAGFIG---------- 235 (to 371)
T.l.MalK 211 AVMNRGVLQQVGSPDEVYDKPANTFVAGFIG---------- 241 (to 372)
GlcV 212 GVLVKGKLVQVGKPEDLYDNPVSIQVASLIG---------- 242 (to 353)
MJ0796 216 IYLKDGEVER-EKLRGFDDR---------- 235 (end)
E.c.MsbA 550 VVVEDGVIVERGTHNDLLEH-RGVYAQLHKMQFGQ--------- 582 (end)
V.C.MsbA 550 LVVDEGEIIERGRHADLLAQ-DGAYAQLHRIQFGE--------- 582 (end)
TAP1 714 LFLEGGAIREGGTHQQLMEK-KGCYWAMVQAPADAPE------- 748 (end)
HlyB 209 IVMEKGKIVEQGKHKELLSEPESLYSYLYQLQSD---------- 241 (end)
HisP 224 IFLHQGKIEEEGDPEQVFGNPQSPRLQQFLKGSLKKLEH----- 262 (end)
MJ1267 224 YVMFNGQIIAEGRGEEEIKNVLSDPKVVEIYIGE---------- 257 (end)
BtuD 204 WLLKGGKMLASGRREEVLTPPNLAQAYGMNFRRLDIEGHRMLIS 247 (end)

Figure 4- Sequence alignment of several ABC ATPases whose structures have been determined. The sequences are ordered based on their homology to the maltose transporter, MalK. Some of the conserved sequence motifs are highlighted. Secondary structure elements for MalK are shown above the sequence. Figure adapted from [72].

The NBDs include several highly conserved short motifs (see Figure 4 and Table 1) associated with nucleotide binding and hydrolysis, such as the Walker [8, 73, 74] and the ABC signature motifs [72]. Mutations in these regions often severely reduce or eliminate ATPase activity and allocrite transport [72, 75]. However, it is important to refer that the sequence identity between the NBDs is far more extensive than the nucleotide binding motifs and spans over the entire catalytic domain.

The X-ray structures of several NBDs, either isolated or in the dimeric form, from different ABC members (including both importers, such as HisP, GlcV, MJ1267, MalK, as well as exporters, like TAP, HlyB, MJ0796), have been reported [33-56]. The structures, like the sequences, are all very similar to each other, and no clear distinction can be made between importer and exporter ABC NBDs. Due to the sequence and structural similarities of these domains, it is reasonable to expect similar mechanisms for the coupling of ATP hydrolysis to an appropriate biological event.

The various NBD dimer X-ray structures available [76] allowed the determination of the correct dimer interface for the NBDs and explained why all ABC transporters require two ABC components for proper functioning (see Figure 5).

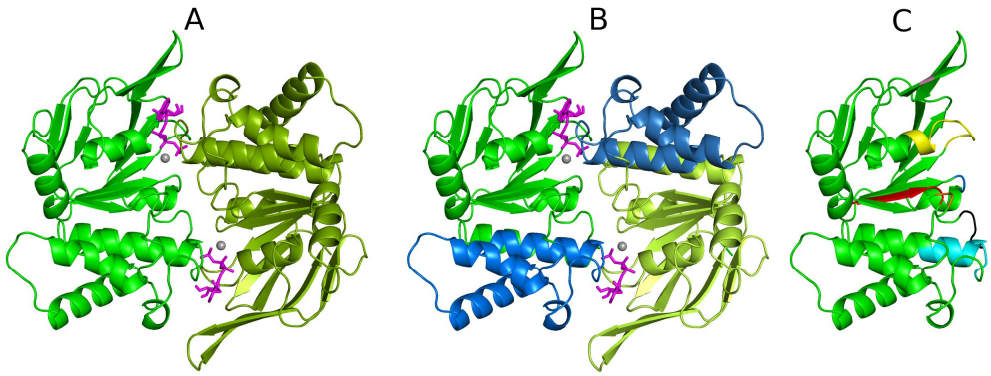


Figure 5- Structure of an ABC transporter NBD dimer. **A**-The NBDs are colored in green (NBD_1 in dark green and NBD_2 in lighter green) and represent the closed dimer conformation, with ATP, (colored in magenta) bound at the interface. The magnesium ions are represented as gray spheres located in the nucleotide binding sites. **B**- Each NBD monomer has a bilobal design and is composed by two distinct sub-domains (the RecA-like and the helical sub-domain). The RecA sub-domain is colored in green while the helical sub-domain in blue. **C**- Schematic representation of NBD_1 with the ATP-binding family motifs highlighted in different colors. The residues and motifs coordinating the nucleotide are colored as follows: A-loop (pink), P-loop (yellow), Q-loop (orange), ABC Signature Motif (cyan), D-loop (black), Walker B (red) and the H-motif (blue).

Each NBD monomer consists in two sub-domains (Figure 5B), a larger one similar to the core structure found in many RecA-like ATPases, and a smaller one, predominantly helical and that is unique to the ABC family (named helical sub-domain). The RecA-like sub-domain usually consists of six α-helices and two β-sheets, and includes three conserved sequences (see Table 1), the Walker-A, and the Walker-B motifs (which are part of the nucleotide binding site) and the A-loop. The Walker-A (GxxGxGKS/T, where x is any amino-acid) and the Walker-B motif (ϕϕϕϕD, where ϕ is a hydrophobic residue) are not unique to the ABC transporters, and are present in many other nucleotide binding proteins. These motifs are essential to nucleotide binding (especially to phosphate binding) and to magnesium coordination. The Walker-A motif (also named P-loop) follows β-strand 3, and forms a loop that binds to the nucleotide phosphates (Figure 5C and Table 1), while the

Walker-B motif is located in β-strand 7 and its terminal acidic residues coordinate the magnesium ion through a H_2O molecule [35, 36, 38]. Additionally, the A-loop [77] is located just above the adenine ring of the nucleotide and contains a highly conserved aromatic residue, which is involved in π–π stacking with the nucleotides.

Table 1- Conserved sequence motifs in the NBD and its functions. Table adapted from [72].

Motif	Consensus sequence	Function	Representative Structures[a]
Walker A or P-loop	GxxGxGKS/T[b]	ATP binding	HisP, MJ076, MJ1267, E.c.MalK, Rad50
Q-loop or lid	Q	a) Interdomain interaction b) H-bond with Mg c) Binding to the attacking water	a) BtuCD b) MJ0796(E171Q), GlcV(ADP) c) MJ0796 (E171Q)
LSGGQ or ABC-motif	LSGGQxQR[b]	ATP binding	MJ0796 (E171Q), E.c.MalK
Walker B	ϕϕϕϕD[b]	contact with Mg mediated by a water molecule	GlcV, MJ1267 (ADP), MJ0796 (ADP)
	E (next to Walker B)	a) Binds to attacking water b) Mg water bridged contact	a) MJ0796 (ADP) b) GlcV
H motif or switch region	H	Binds to the γ-phosphate	MJ0796 (E171Q), E.c.MalK
A-loop	A	Binds to the Adenine ring of the nucleotide	MJ0796 (E171Q)

[a] References for some exemplifying structures: HisP [33], MJ0796 [36], MJ1267 [35], Rad50 [78], TAP1 [34], GlcV [38], MalK [52], BtuCD [57], MJ0796 (E171Q) [51].

[b] x represents any amino-acid, ϕ represents a hydrophobic and A an aromatic amino-acid.

Lastly, another very important residue located in the RecA sub-domain, is the acidic residue (a glutamate in many cases) immediately following the Walker-B motif, which is thought to act as a general base in hydrolysis ("general base" model) [79, 80]. This residue is known to interact with the magnesium ion [51, 80] and in several of the X-ray structures available (e.g. [51, 80]), it makes a hydrogen bond with the alleged hydrolytic water in the binding site. Furthermore, the mutation of this acidic residue by its corresponding amide completely abolishes, in some cases, the ATPase activity [46, 80, 81] and triggers the formation of a trapped ATP-bound dimer (as for example in [51, 80]). These observations led to the hypothesis that this residue is capable of polarizing the attacking water molecule for hydrolysis [79, 80] and that the corresponding amide residue is not able to act as a base for hydrolysis.

The above described "general base" model is not consensually accepted and there is experimental data that does not support the previously described glutamate function. For example, in some cases, such as P-glycoprotein, HlyB, and GlcV, the glutamate to glutamine mutation does not completely abolish the ATPase activity [38, 54, 82, 83], a fact that can not be explained by the "general base" mechanism. Therefore, an alternative model, called "substrate-assisted catalysis", was proposed, based on the results obtained for the haemolysin transporter [54]. In the "substrate-assisted catalysis" model, the Walker-B glutamate residue, together with the H-loop histidine, forms the catalytic dyad essential for hydrolysis. The histidine is thought to work as a "linchpin" holding together the ATP γ-phosphate, the attacking water molecule, the magnesium ion, and other catalytically important amino acids, while the Walker-B acidic residue only restricts the histidine flexibility in order to adopt a catalytically competent conformation [6, 54, 84]. Nevertheless, this model is also not bulletproof and it can not explain why, while in the haemolysin transporter the substitution of this histidine residue completely eliminates hydrolysis [54], and in the intact maltose transporter, the histidine to arginine mutation retained some residual activity (2% of the wild-type ATPase activity) [85].

Additionally to the RecA-like sub-domain, each NBD monomer is also constituted by a smaller sub-domain (called helical sub-domain), which is formed by three or four helices and by the ABC-signature motif (which is a sequence unique for the ABC family and distinguishes ABC proteins from other nucleotide-binding proteins). The ABC- signature motif (see Table 1) binds the ATP γ-phosphate and this, together with the P-loop and the Walker-B motif, forms the nucleotide binding site [72].

The Rec-A and the helical sub-domains are joined together by two flexible loops, one of which contain a highly conserved glutamine residue (called Q-loop) and the other a conserved histidine residue (named H-loop). The Q-loop (also named “the lid” [50] or the “γ-phosphate switch” [36]) is located in β-strand 6 and contains a glutamine residue that binds to the magnesiun ion and to the attacking water [38, 51] (see Table 1). Based in the high-B factor values reported in several of the NBD X-ray structures, the Q-loop region appears to be highly flexible when ATP is not present. Furthermore, over the last years, several authors have suggested that the Q-loop glutamine residue plays a central role in mediating the allostery between the nucleotide binding site and the helical sub-domain [51, 86, 87]. The potential relation between this glutamine residue and the helical sub-domain will be extensively addressed later in section 3 and 4. The H-motif or the “switch” [50] (see Table 1), is located in β-strand 8, and contains a highly conserved histidine residue that forms hydrogen bonds with the γ-phosphate of ATP [51, 52].

The NBD physiological dimer is formed by two monomers in a head to tail arrangement (see Figure 5A), with the two nucleotide molecules bound at the interface of the monomers, interacting with both residues from the P-loop of one monomer and the ABC-motif of the other. The sharing of nucleotide-binding sites between the monomers explains the need to have two NBDs. Moreover, the

observation of positive cooperativity in hydrolysis reported for some members (such as MalK [88], YvcC [89] and HlyB [90]) can be easily understood.

Over the last years, several high-resolution structures for the NBD monomers and dimers have been determined [33-43, 47, 50-54, 91, 92], allowing us to gain relevant insights into the working mechanisms of these nanomachines. In order to obtain crystals for the pre-hydrolysis dimer (with ATP-bound in the binding sites), some residues in the active site were mutated [46, 51, 54] or EDTA was used to avoid the ATP hydrolysis [52]. Based in the ATP-bound crystal structures available, a complex network of interactions between the protein and the nucleotides is now unraveled. First of all, the nucleotide adenine ring is normally stabilized by π-π stacking interactions with the A-loop conserved aromatic residue [77, 79]. Moreover, the P-loop conserved lysine residue forms hydrogen bonds with the oxygen atoms of the α- and γ- phosphates, thereby holding them in a defined orientation, whereas the magnesium ion is coordinated by the oxygen atoms of the β- and γ-phosphates and by some P-loop residues [34, 38]. Additionally, the highly conserved H-loop histidine forms a hydrogen bond with the γ-phosphate, which is known to be essential for hydrolysis [52 , 54]. The side chain of the serine and the backbone amide groups of the ABC signature glycine residues are known to coordinate the γ-phosphate in ATP.

1.2.2.1. NBDs asymmetry, one of the oldest unanswered question

Although the two NBD monomers are strictly necessary to hydrolysis, the question of their equivalence on binding sites symmetry is still open, despite all the experimental and theoretical data available. Of course that, for the cases were the NBD dimer is a heterodimer (with the two monomers displaying different amino-acid sequences), as for example in CFTR, the asymmetrical behavior of the monomers and the binding sites is easily explainable. However, over the last years, several experimental and theoretical works (performed whether in the isolated ATPase

dimer, e.g. [1, 47, 87, 93-98] or in full length transporters, e.g. [99-101]) pointed out to the fact that a functional asymmetry can be also observed in NBDs homodimers, such as HlyB [47], histidine [99] and maltose transporter [100, 101]. For example, in the histidine transporter [99], it has been reported that thiol-specific reagents react more readily with a cysteine residue from one monomer than with the equivalent cysteine in the other. In the maltose transporter, it has been reported that the cross-linking pattern between cysteines located in the helical sub-domains and in the TMDs is different for the two monomers [100, 101]. Nevertheless, in some cases, like in the maltose transporter, the NBD dimer asymmetric behavior reported can easily be explained by the effect exerted by the TMDs in the NBDs, since the former can have different amino acid residue sequences. Moreover, in the case of importers, the existence of a substrate binding protein (SBP) may introduce an extra source of asymmetry, since the SBPs are themselves asymmetric [102] and by this reason imposing functional asymmetries in the NBD monomers.

Despite all the information available until this moment, the question whether the two NBD monomers and the two binding sites are functional equivalents, does not have a straightforward answer and it may depend on the transporter being homo- or heterodimeric (both for the TMDs and for the NBDs) or if it requires a SBP. However, there are some experimental evidences (such as [47, 100, 101, 103-105]), which suggest that functional asymmetries in the binding sites can arise during the catalytic cycle. These experimental observations lead to the formulation of the hypothesis that the observed NBD asymmetry is an intrinsic and essential characteristic of an ABC pump and the structural basis for the ATP hydrolysis mechanism.

The question of the NBD monomers asymmetric behavior will be widely exemplified and debated throughout this thesis, in sections 3, 4 and 5.

1.2.2.2. Stoichiometry of ATP binding/hydrolysis and the role of two binding sites

Two other still unanswered questions associated with the ATPase domains are related with the role of the two binding sites in transport, and with the number of ATP molecules hydrolyzed per transport cycle. Since the nucleotides species are bound to the NBD monomers interface, the existence of positive cooperativity between the nucleotide binding sites (as observed for example for the maltose and histidine transporters [88, 106]) is currently easily explained. However, what seems to be unclear until today is whether hydrolysis in one or in the two binding sites is necessary to complete a single transport cycle. There are some examples (such as the ribose transporter) where only one of binding sites posses all of the highly conserved residues considered essential for hydrolysis [107], suggesting, in this case, that hydrolysis in one binding site only is enough for allocrite transport. Supporting this hypothesis, the mutation, in the intact histidine transporter, of the catalytically histidine residue in one of the two NBD monomers (while the second site remained catalytically competent) seemed to be relatively well tolerated and not sufficient to eliminate transport [108]. However, and contradictorily, a similar substitution in a single binding site, in the maltose transporter, severely impaired the ATPase activity and transport, suggesting in this case, that hydrolysis at both sites is essential for adequate functioning [85]. Additionally, the mutation of the possible catalytic glutamate (the acidic residue located immediately following the Walker-B motif), when performed in a single site of the complete P-glycoprotein transporter, completely inactivates the transporter [82], but when performed in the isolated NBDs dimers of, for example GlcV and HlyB [39, 90], allows the retention of some ATPase activity. Nevertheless, in the last cases (GlcV and HlyB), the apparent contradictory observations may be explained by the loss of functionally important cooperative interactions between the NBDs and the TMDs, in the experiments where the TMDS are stripped out [7].

Over the last years, many efforts have been made in order to measure the stoichiometric ratio of substrate transport and ATP hydrolysis, both in vivo and in vitro, for several ABC members, hoping to finally determine how many ATP molecules are hydrolyzed per transport cycle. Experimental technical details make these measurements very difficult to achieve, due, especially, to the high levels of uncoupled ATP hydrolysis and the possible leakage of substrate from membrane vesicles. Moreover, when the stoichiometric ratio can be effectively measured, most of the times, the values reported are so disparate that is impossible to conclude anything, like for example in the maltose transporter (where rates ranging from 1 [109, 110] to 17 [111] ATPs per transported sugar have been reported).

After this brief description of the ABC catalytic domains chemistry and despite the available information (many times contradictory), the reader clearly understand that we are still very far away from a complete understanding of the ABC molecular mechanisms. Nevertheless, the fact that the ABC modules are highly conserved (both in sequence and in structure) between all ABC members raises the hypothesis that the basic mechanisms for nucleotide hydrolysis and their coupling to allocrite transport, are probably conserved throughout the ABC transporter family.

1.2.3. Substrate Binding Protein (SBP)

Most ABC importers rely on the presence of soluble high-affinity extra-cytoplasmic binding proteins (SBPs) for proper functioning. Historically, SBPs were the first components of ABC transporters to be identified [112]. Several experimental studies demonstrated that these proteins are absolutely required for the proper functioning of the import system to which they are associated [113, 114]. SBP are essential, even at high substrate concentrations and, for example, in the *E. coli* maltose importer system, the deletion of the gene encoding the maltose binding protein severely impairs transport [113].

In Gram-negative bacteria, SBPs are located in the periplasmatic space between the inner and outer membranes whereas in Gram-positive organisms and in Archea (which lack the periplasm), SBPs are often lipoproteins bound to the external face of the membrane or directly fused to the membrane transporter itself [115].

The range of allocrites retrieved by the SBPs is extremely diverse [7] and includes mono- and oligosaccharides, organic and inorganic ions, amino acids and short peptides, iron-siderophores, metals, polyamine cations opines, and vitamins. Moreover, most SBPs are specific for a single allocrite or for a family of structurally related substrates [7]. However, there exceptions to this rule, when a single SBP is very versatile and is able to handle structurally unrelated substrates [7].

SBPs are usually monomeric and both structural and kinetic data indicate that there is only one substrate-binding site per SBP [116]. Moreover, the X-ray structures of several full length transporters (MalFGK$_2$E [62], ModABC [60] and BtuFCD [61]) in complex to their SBPs revealed that only one SBP directly interacts with the transporter membrane domains at each time. However, in order to increase the transporter versatility, there are cases where multiple SBPs, with different binding specificity, are able to interact with the same transporter (like in the lysine/histidine/arginine transporter [117]). Until this moment, most of the SBP knowledge arises from the study of Gram-negative bacteria, as only a few Gram-positive and Archaea substrate-binding lipoproteins have been characterized, either biochemically or structurally [118-120].

The available crystallographic structures of SBPs [91, 102, 121-125] show that all SBP proteins adopt a relatively similar fold formed by two globular domains or lobes (called the N- and C-lobes, since they contain the protein's N- and C- terminals). These lobes are connected by one or more linker segments, and the substrate binds in between these two domains [102] (Venus's Flytrap model [126]). Each lobe

is composed of an alpha-beta fold consisting of several β-sheets surrounded by α-helices and connected by loops. Over the years, high-resolution structures of many periplasmic SBPs in multiple conformations (open-unliganded, closed-unliganded, open-liganded, and closed-liganded) conformations have been determined [91, 102, 121-124, 125] which allows us to understand the conformational changes induced upon substrate binding (see Figure 6). Based on all X-ray structures, three distinct folding patterns have been identified for the SBPs, which allow the division of SBPs in three main groups [121, 127]. The classes I and II [72], represented by the simple sugar (arabinose and ribose) binding protein and the maltose/maltodextrin binding protein [128], respectively, differ primarily in the fold of the two domains. In class I, the β-sheet topology for the N- and the C-lobes is $\beta_2\beta_1\beta_3\beta_4\beta_5$, whereas class II has a different topology ($\beta_2\beta_1\beta_3\beta_n\beta_5$ with n representing the first cross-over from the N-terminal to the C-terminal domain). Additionally, in class I, the hinge-region is generally formed by three connecting strands whereas in class II it only has two connecting strands. However, these two classes (I and II), upon substrate binding, undergo a conformational change involving the bending of a hinge that joins the two lobes together. In the substrate-free conformation, the C- and N-lobes are well separated, with an open, solvent-accessible cleft between them. Upon substrate binding, the substrate promotes a substantial domain rotation and closing of the cleft around the substrate [121] (see orange cartoon in Figure 6).

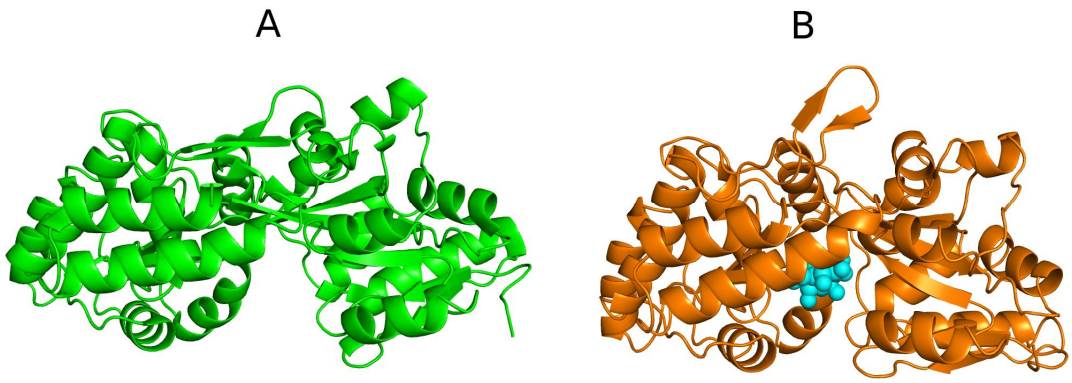


Figure 6- Maltose binding protein (MBP) in two distinct substrate-bound states. **A**- open-free state (PDB entry 1OMP [129]). **B**- closed maltose-bound conformation (PDB entry 1ANF [130]). The substrate, when present, is represented by cyan spheres and it is located in the cleft between the N- and the C-lobe.

Later, a third class of SBPs (class III) was defined and its distinguishing feature is a single α-helix segment linking the two α/β lobes. The class III SBPs, represented by the iron/siderophore/vitamin B_{12} SBP [91, 131], present a more rigid hinge structure and substrate binding is not accompanied by a large domain movement, like the one documented for classes I and II [91, 132].

The surface of the SBP that interacts with the transporter TMDs is, in most cases, already well defined (by mapping the mutations that affect transport [133, 134] and by X-ray crystallography [60-62]) and it involves both SBPs lobes and both hydrophobic transmembrane domains (see Figure 7 for an example).

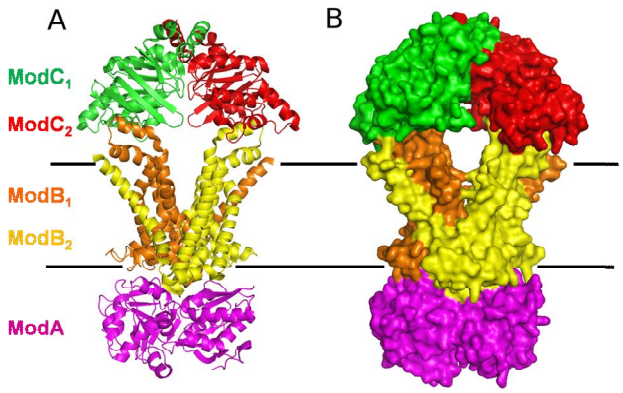


Figure 7- Representation of the complex formed by the putative molybdate importer ($ModB_2C_2$) and its SBP (ModA) [60]. The ModB subunits are colored in orange and yellow whereas the ATPase domains (ModC) are highlighted in green and red. The ModA protein is colored in magenta. The horizontal black lines depict the probable location of the lipid membrane, based on the hydrophobicity of the protein residues. The reader should note that the SBP substrate-binding cleft is more or less located close to the interface between the TMDs.

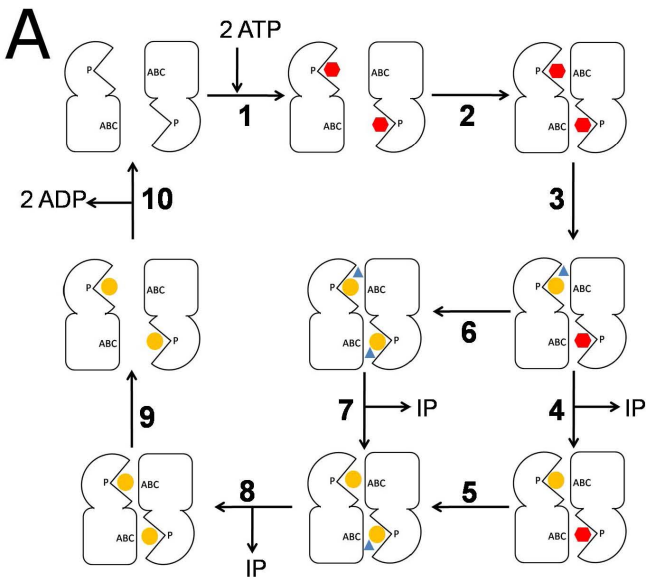
Although the substrate binding components are essential for the function of the transporter to which they are associated, they are not considered part of the minimum functional unit, since many ABC transporters (including all exporters) do not require an equivalent component to achieve transport. Over the last years, questions like how do SBP recognize its substrate or the nature of the substrate induced conformational changes have been extensively addressed. Excellent reviews are available (such as [128] or [127]) for a more detailed description of the advances in the SBP field.

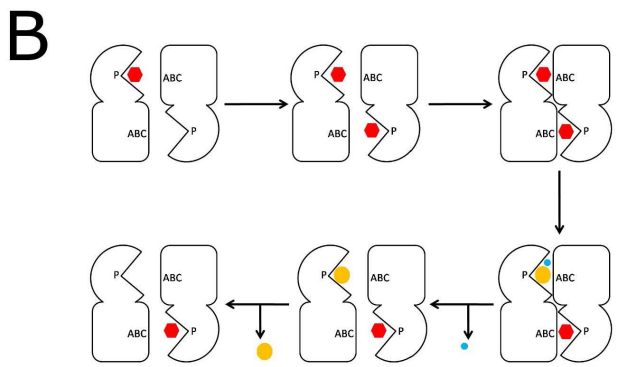
1.3. Molecular Mechanisms in ABC transporters

1.3.1. Mechanisms for NBD dimer functioning

Despite the large amount of experimental data available and all the detailed insights into the structure of the ABC modules, the precise mechanism for NBD dimer functioning in ABC transporters is still controversial. However, based on the NBDs highly conserved sequences and fold, a common and unique mechanism for ATP hydrolysis is likely to exist, independently of the nature of the transported allocrite. Until this moment, three models have been proposed for the ABC ATPase domains functioning [135]: the “processive-clamp”, the “alternating two sites” and the “constant-contact” model. The major differences between these three models are related to the number of ATP molecules required to open the dimer interface and the extension of the dimer interface opening. In this chapter, the current models for NBD dimer formation coupled to ATP hydrolysis will be briefly described.

One of oldest questions related to hydrolysis (in the NBDs binding sites) is whether it occurs in a sequential or in an alternating mode. This point is the main difference between the “processive-clamp” (Figure 8A) and other two models, the “alternating two site” (Figure 8B) and the “constant-contact” model (Figure 8C).


A
2 ATP
1
2
3
4
IP
5
6
7
IP
8
IP
9
2 ADP
10
ABC
P


B
ABC
P

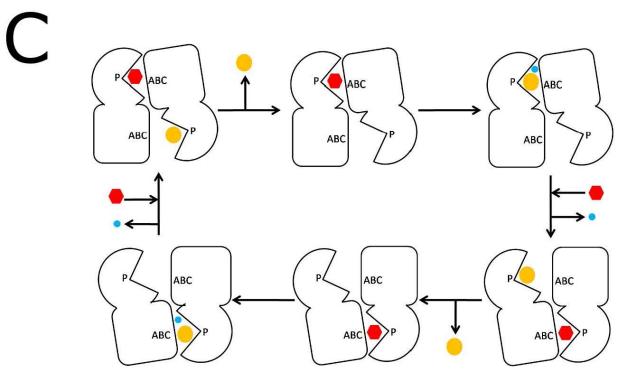

C
ABC
P

Figure 8- Mechanistic models of dimer formation coupled to ATP hydrolysis in the NBDs of an ABC transporter. ATP is represented as a red hexagon, whereas ADP and IP are depicted by yellow and blue spheres. Each NBD monomer is formed by two sub-domains (RecA-like and helical sub-domains). The letters P and ABC refer to the P-loop and to the ABC signature motif. **A**- "Processive-clamp" model [51, 52, 136-138]. **B**- "Alternating two sites" model [6, 139]. **C**- "Constant-contact" model [87, 97, 140].

The "processive-clamp" [51, 136], "tweezers-like" [52] or "switch" model [137, 138]. (Figure 8A) was proposed based on experiments performed in the Md1P transporter [136]. For the Md1P transporter, it was possible to isolate three different dimeric intermediate states [136], each representing different steps in the catalytic cycle; two ATP-bound, one ATP- one ADP- bound, and two ADP-bound states. With this work it was possible to show that ATP binding to the two binding sites induces dimer formation. Additionally, it was also observed that two nucleotides are always present in the binding sites, the inorganic phosphate is able to dissociate from the dimer and, finally hydrolysis occurs in a sequential mode. In the "processive-clamp" model (Figure 8A), the binding of a pair of ATP molecules (step 1) to the spatially separated NBD dimer induces the formation of the functional closed dimer (step 2), which is then followed by the hydrolysis of one of the two ATP molecules (step 3). After this first hydrolysis, the inorganic phosphate (IP) is either released first (step 4), followed by the hydrolysis of the other ATP (step 5) or the second ATP is hydrolyzed first (step 5) and then the two IP molecules formed are released (steps 7 and 8). Upon IP exit, the dimer disassembles (step 9) and the ADPs are set free (step 10), restarting the cycle all over again. As can be seen from Figure 8A, in this model the IP exit can occur prior to ADP release and dimer dissociation. However, how the IP release can be achieve is not yet known, but some possible exit ways have been proposed, as for example the opening of hydrophilic tunnels providing a continuous passage from the nucleotide binding site to the surface [47] or a gate opening (after hydrolysis) which exposes the IP species to the solvent [1]. At this moment, it is becoming clear that IP exit

mechanisms are not common to all NBDs, and alternative mechanisms can appear for other ABC members (for a more intensive discussion in this subject we suggest reading Section 3.4.3).

It is noteworthy that the majority of the experimental data supporting the existence of symmetrical sandwich dimer states (such as the two ATP- ot two ADP-bound states) comes from structural and biochemical studies performed in the absence of the catalytic magnesium ion or in ATPase inactive mutants, or using non-hydrolysable ATP analogues [135, 141]. Additionally, there are several X-ray structures in which the NBD monomers are separated more than 20 Å (e.g. [52, 65, 66, 142]). These high distances values observed for several nucleotide-free structures, together with the fact that ADP and ATP differ by only one charge unit, lead some researchers that claimed to be unlikely that electrostatic forces could bring the monomers together at all, or at least in a biological relevant timescale [135]. Due to all these questions, the earlier version of the “processive-clamp” model [138] was recently updated to a new version [143], where the authors propose that structural differences between the open and closed dimer conformation are probably less extended than the complete physical separation of the dimer.

In alternative to the “processive clamp” model, a second model for hydrolysis, called the “alternating two sites” or “two cylinder engine” model (see Figure 8B), was proposed by Senior and coworkers [6, 139] based on their work with P-glycoprotein. This model is based on the experimental observation that vanadate can replace the IP species formed during ATP hydrolysis [139, 144], mimicking the step where both post-hydrolysis products are simultaneously present in the binding site. Moreover, the observation that the trapping occurs in just one of the two binding sites, suggested that hydrolysis can take place in an alternated way (since in P-glycoprotein both sites are catalytically competent and can hydrolyze ATP), and only one ATP molecule is hydrolyzed each time an allocrite is transported [139,

145]. Additionally, it is implicit in this model that the transporter retains memory of which site last hydrolyzed ATP, so that the next hydrolysis would occur in the opposite site, resulting in an alternating ATP hydrolysis [6, 139]. In this model (see Figure 8B), in the ATP-bound closed dimer, only one of the nucleotides is hydrolyzed before the respective monomer opens up. ADP is then exchanged and the binding of a new ATP molecule can then occur, thus inducing the formation of a closed dimer again. After the first hydrolysis and nucleotide exchange in one of the monomers, the ATP in the second monomer will then be hydrolyzed. However, this model, although derived from biochemical studies, is incompatible with the X-ray structures available, since all structures unambiguously show symmetric bound nucleotides states (with two ATP or ADP-bound) and, until this moment, no hybrid nucleotide bound state (e.g one ATP and one ADP-bound) was reported.

A third alternative model, the “constant-contact” model [87, 97, 140] (Figure 8C) was proposed, also based on an alternating sites scheme [139]. In this model, and although the binding sites hydrolyzed ATP and open in an alternated way, the two NBDs always remain in close contact in the opposite unopened binding site [86, 140]. Since the NBDs are always in straight contact throughout the catalytic cycle, the opening and closing of the binding sites occur via intra-subunit conformational changes within the NBDs dimer. Until this moment, a lot of experimental data (such as enzymological data, EPR analysis or cross-linking experiments of cysteines) supports the asymmetric model for NBD function [141, 146-149], and the hypothesis that the NBDs do not dissociate and move apart substantially during the catalytic cycle [146, 150]. Recently, the “constant-contact” model gained additional support from several theoretical works [2, 3, 87, 96, 97] that were able to show that the opening of the dimer interface (upon ATP hydrolysis) occurs in only one of the two binding sites via the alternated rotation of the RecA-like sub-domain relative to the helical sub-domain combined with the loss of the interactions between the ABC motif and the nucleotide. Nevertheless, this model is incompatible with the X-ray structures available, since the nucleotide-free structures clearly show that, in the

resting state, the NBDs are dissociated and separated by, in some cases, distances up to 20 Å (e.g. [52, 65, 66]).

Despite all three previous models being corroborated by experimental data, it is important to refer that, until now, no single experiment was able to convincingly rule out any one of them.

One essential task for ABC transporters is the harnessing of the energy released from nucleotide binding/hydrolysis and its conversion into mechanical work (most probably through conformational changes). Consequently, one of the major questions still awaiting for an answer is which are the conformational rearrangements induced by nucleotide binding and hydrolysis in the NBDs, and how are these structural changes transferred to the TMDs, ultimately allowing allocrite transport. Upon comparison of the different nucleotide-bound NBDs monomeric structures (e.g. [35-37]), the major conformational change observed between different nucleotide-bound states is the rigid-body rotation of the helical sub-domain relative to the RecA domain [7, 72], a movement which has been suggested to facilitate the nucleotide exchange upon hydrolysis [35].

Moreover, and since the ABC modules function as a dimer, it is important to analyze the conformational changes in the context of the dimeric structures. The nature of the conformational changes though to occur can be illustrated by the crystal structures of the maltose transporter NBD dimer (MalK). The MalK dimer was crystallized in three distinct nucleotide-bound states: the resting-, the ATP-bound, and the ADP-bound states [52, 53]. The crystallization was only possible due to the presence of an extra regulatory domain, located in the MalK C-terminal region, which stabilizes the MalK dimer in the absence of the TMDs and the nucleotide species [151]. The different MalK crystal structures (see Figure 9) clearly shown that the two monomers are kept together during the cycle by the RDs, and that the NBD region open and closes like the tips of a pair of tweezers.

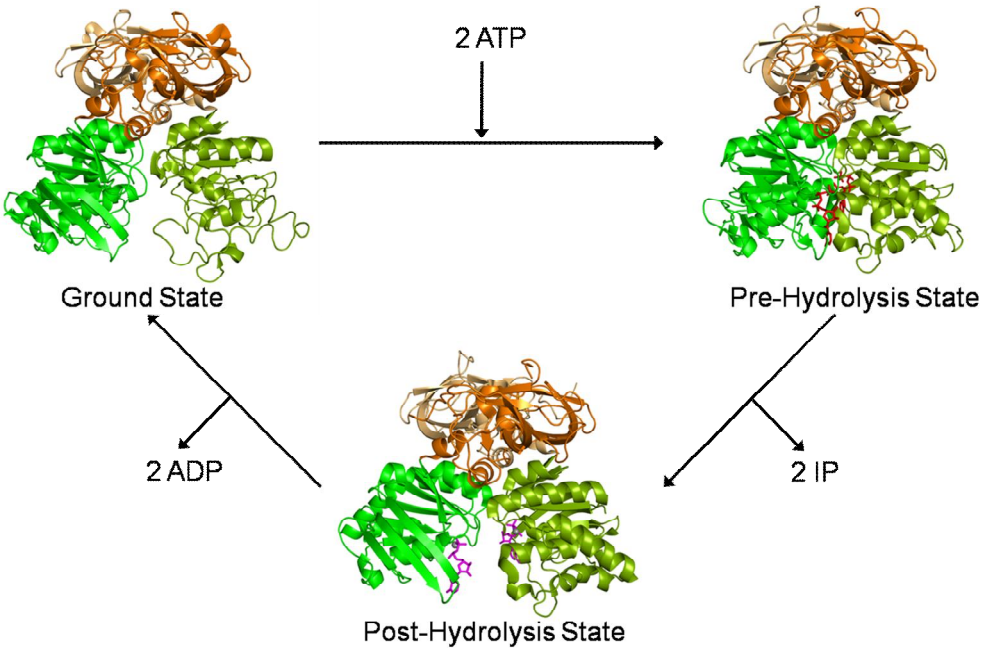


Figure 9- Structures of the MalK dimer in the nucleotide-free [52, 53], ATP- [52, 53] and ADP-bound states [52, 53]. In the absence of nucleotides (ground state), the two MalK monomers are separated from each other, held as a dimer primarily through contacts between the C-terminal RDs. In the presence of ATP (pre-hydrolysis state), the NBDs are closed, allowing ATP hydrolysis to occur. In the ADP-bound state (post-hydrolysis state), the NBDs are again separated in a conformation similar to the resting state. In this figure the NBDs are colored in green while the RDs are colored in orange. Corresponding domains in the second MalK monomer follow the same color scheme but are rendered in lighter colors. ATP and ADP are represented as red and magenta sticks, respectively, and magnesium as a gray sphere.

As can be seen in Figure 9, in the nucleotide-free structure (ground state), the MalK N-terminal regions are separated and the dimer is maintained solely through contacts in the C-terminal RDs. Upon ATP binding, the NBDs get closer, closing the dimer with the two ATP molecules buried along the interface. After ATP hydrolysis, the two NBDs separate from each other inducing the dimer dissociation and the release of the post-hydrolysis products. The comparison between the pre-

hydrolysis and the pos-hydrolysis structures led to the conclusion that ADP, unlike ATP, cannot stabilize the closed NBD dimer in this transporter [49]. The main rearrangements between the three different conformations are achieved, firstly, by a rotation of the entire NBD relative to the RD and, secondly, by a rigid-body movement between the RecA and the helical sub-domains, which results in an inward movement of the helical sub-domain towards the nucleotide binding site of the same monomer. In contrast to the NBDs, the interactions between the RDs are essentially unchanged during the catalytic cycle.

1.3.2. Mechanisms for coupling hydrolysis to allocrite transport

Several experimental studies have early showed that the NBDs bind and hydrolyze ATP and that nucleotide hydrolysis is coupled to transport [74, 110, 152, 153]. Nevertheless, how ATP hydrolysis is coupled to allocrite translocation is a question still waiting to be answered. Additionally, the conformational changes (both in the NBDs and the TMDs) responsible for vectorial transport in ABC transporters are also mostly unknown.

There are several models proposed for coupling ATP hydrolysis, whether to exporters (Figure 10 and Figure 11) or to SBP-dependent importers (Figure 12). These models differ in a number of aspects, including the number of hydrolyzed ATP molecules per transport cycle, the number of allocrite binding sites and the point in the transport cycle when the allocrite is translocated. In the following sub-chapters, these proposed transport models will be briefly described.

1.3.2.1. Export Systems

At least four models for coupling hydrolysis to transport in ABC exporters have been reported in literature (Figure 10 and Figure 11). These models differ mainly in the number of allocrite-binding sites in the TMDs, the transport power stroke and in the number of ATP molecules hydrolyzed per transported allocrite.

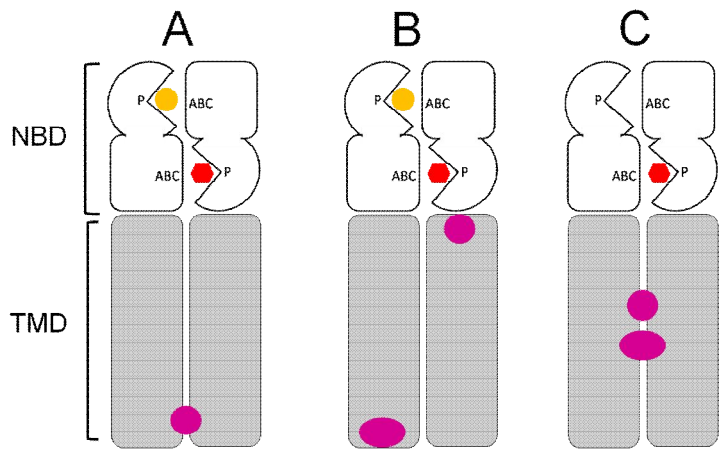


Figure 10- Three of the four possible models for exporter pumps. In this figure, ATP and ADP are represented as red hexagons and yellow spheres, respectively (**A**) In this model [139], the magenta sphere represents the single allocrite binding site shared between the two TMDs (**B**) In this model [154], the magenta sphere represents the allocrite binding site inward-facing conformation whereas the magenta ellipse represents the outward-facing conformation. (**C**) In this model, two binding sites coexist [155], an inward-facing site (magenta sphere) and an outward-facing site (magenta ellipse).

The model A (Figure 10A), called the “alternating catalytic sites” model, was proposed by Senior and co-workers [139] based on observations made on the mammalian P-glycoprotein. In this model, only one interchangeable drug-binding site is shared between the two halves of the pump, and the conversion from an inside-facing, higher affinity, to an outside-facing, lower-affinity conformation, is associated with one ATP hydrolysis [139].

The model B (Figure 10B), which is a variation of the “alternating catalytic sites” model [139], was proposed by van Veen and co-workers [154] based on their work with LmrA. In this model, each half-transporter has its own drug binding site and while hydrolysis and transport occurs in one half of the transporter, the other half resets its drug-binding site from the outward-facing, low-affinity conformation, to the inward-facing, high-affinity conformation [154].

The model C (Figure 10C) was proposed by Sauna & Ambudkar [155] based also in the mammalian P-glycoprotein. In this model, both high- and low-affinity sites coexist, and during the transport cycle, the loss of high-affinity binding promotes the transfer of the allocrite from the high-affinity to the low-affinity site [155]. While in the first two models, drug transport occurs each time ATP is hydrolyzed, the third model suggests that ATP hydrolysis at the second site is required to reset the system before transport can occur again [155].

A fourth model for transport (see Figure 11), the “ATP-switch” model, was proposed by Higgins and colleagues [138, 156], based on their work with the mammalian P-glycoprotein. In this model, the authors suggest that nucleotide binding to the transporter, rather than ATP hydrolysis, drives the major conformational changes that reorient the drug-binding site from the inside, high-affinity to the outside, low-affinity conformation [156].

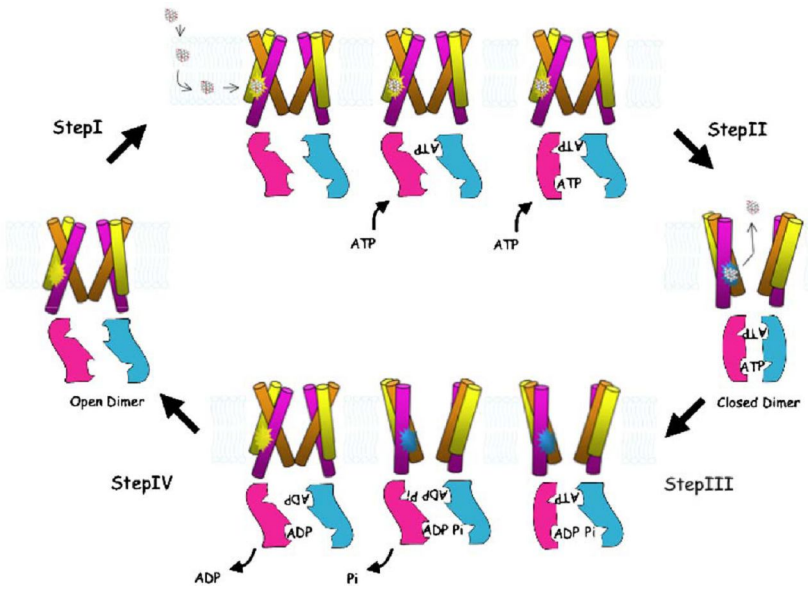


Figure 11- The ‘ATP switch’ model for the transport cycle of an ABC transporter [138, 156]. The represented scheme is for an exporter and some variations may be required for importers. The TMDs are represented as yellow and purple ‘cylinders’ spanning the membrane, while the NBDs are shown

as blues and pink shapes. The yellow and blue stars, located in the TMDs region, represent the high affinity and the low affinity allocrite-binding site, respectively. Figure adapted from [138].

In the “ATP-switch” model [138, 156], transport is a multistep process involving communication via conformational changes, in both directions, between the NBDs and TMDs. In this model, the driver for transport is the switch between two principal NBD conformations: the “closed dimer” (with the two monomers in straight contact and two ATP molecules bound at the interface) and the “open dimer” (where the monomers are separated). The conversion from the “open” to the “closed” NBD dimer conformation induces the conformational changes in the TMDs necessary for transport. The conversion from the “closed” to the “open dimer” conformation happens after ATP hydrolysis and resets the transporter ready for the next transport cycle. The precise conformational changes that trigger the ‘closed dimer’ formation are yet unclear, but experimental studies in intact transporters suggest that the conformational changes at the NBDs are relatively small [157, 158].

1.3.2.2. SBP-dependent import systems

Since all ABC exporters are able to function without resorting to auxiliary allocrite rescuing proteins, SBPs are sometimes considered accessory proteins that give bacteria the ability to transport nutrients with a high affinity. However, they are an important part of the ABC importer translocation machinery and are essential, not only for allocrite uptake but also for coupling transport to ATP hydrolysis.

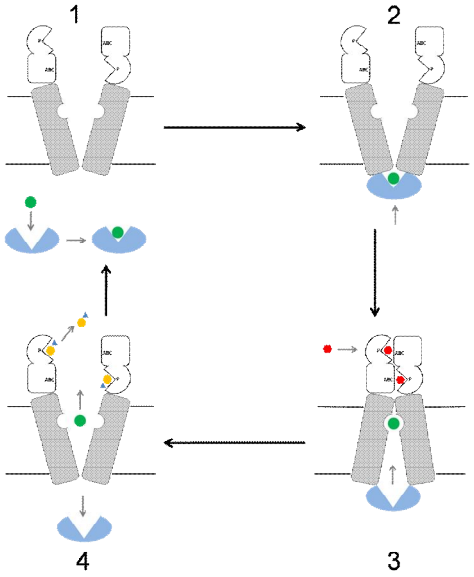


Figure 12- SBP-dependent translocation mechanism for an ABC importer. This model is composed by four distinct steps. For a detailed explanation, see the following text

The simplest model for the SBP-dependent transport is composed of four basic steps (Figure 12): (1) allocrite binding to the SBP, (2) docking of the closed, ligand-bound SBP onto the TMDs, inducing cooperative ATP binding to the NBDs, (3) allocrite transfer from the SPB to the TMDs binding site and reorientation of the transporter from an inward- to an outward-facing conformation, and (4) dissociation of the NBD dimer (after hydrolysis) and allocrite release from the TMDs to the cytoplasm, followed by the transporter reset to its initial position. Alternatively, it has not yet been ruled out that, upon ATP binding, the allocrite is directly transferred to the cytoplasm (step 3) and ATP hydrolysis, in combination with the dissociation of the post-hydrolysis products, is solely used to reset the importer for another translocation cycle.

Nowadays, the essential role played by the SBP in the transport cycle is clearly accepted, and it was shown that the interaction of the allocrite-bounded SBP with the TMDs truly initiates this cycle [159]. It is also known that allocrite analogues failing to induce SBP closure do not stimulate ATPase activity and are not transported, a fact that suggests that the SBP closed conformation is necessary to initiate a cycle of transport and hydrolysis [160].

Although ABC importers and exporters probably share a common translocation mechanism, we can never forget that important structural differences between ABC importers and exporters exist, which may introduce differences in the transport mechanism. However, the significance of these differences in terms of a general ABC mechanism awaits further experimentation, and more biochemical and structural analyses are needed before we can clarify the existence of a global translocation mechanism.

1.4. Computer simulations of ABC transporters

Computational theoretical techniques, such as homology modeling and molecular mechanics/dynamics simulations, are nowadays considered important and powerful tools that can contribute to increase our understanding of protein structure, function and dynamics. In particular for membrane proteins, which are difficult to crystallize due to technical limitations, computational methods may be useful tools for complementing the experimental structural information and providing properties that are hard to be measured experimentally. In this chapter, we will briefly describe the evolution of the computer simulations field in the ABC transporter family and we will show how computational techniques may help obtain important insights into the ABC mechanisms.

1.4.1. Protein dynamics

When a NMR or X-ray structure (preferentially high-resolution structure) is available, the intrinsic dynamic behavior of that protein can be studied using molecular mechanics/dynamics simulations and normal model analysis. Classic atomistic MD simulations of proteins are limited, until this moment, to several tens of nanoseconds. Although this timescale is too short to study the major conformational changes occurring during a transport cycle (≥ 100ms for a transport event in P-glycoprotein [161]), this technique can be used to study individual states or initial transitions in the transport cycle, and can facilitate the interpretation of the experimental data available.

Over the last decade, molecular mechanics/dynamics simulation and normal model analysis have been applied to several complete ABC transporters [93, 95, 162-170] or to isolated ABC transporters domains [1, 56, 87, 94, 96, 97, 171-173], as the reader will see in the following sub-chapters.

1.4.1.1. Nucleotide binding domains

Over the last decade, the dynamic behavior of the NBDs (both in the isolated and in the dimeric form) has been extensively studied [1, 56, 87, 94, 96, 97, 171-173]. In 2002, the first molecular dynamics simulation study of an ABC transporter domain was published. In this work [173], the authors report a 390 ps simulation for the monomeric NBD of a histidine transporter (HisP), aimed at identifying the regions that may be involved in conformational changes in the NBDs. Since then, multiple, much longer (5-20 ns), simulations of the monomeric HisP have been reported [171]. Using 3 different nucleotide-bound HisP conformations (nucleotide free-, ATP- and MgATP-bound), Campbell and co-workers [171] tried to understand how the NBD responds to the presence of ATP. From the ATP-bound simulation, the authors observed that the presence of ATP was associated to the movement of the helical sub-domain region towards the core RecA-like sub-domain. Curiously, a similar movement was not observable in the MgATP-bound simulation, presumably due to the presence of the Mg ion.

Since the two NBDs are known to function as a dimer [174], simulations of the NBD dimer were also performed in order to understand the conformational changes occurring during the ATP-hydrolysis cycle and to determine the relevance of this rearrangements in the NBD:TMD communication mechanisms [87, 172]. In 2005, Campbell and Sansom [172] reported a 10 ns simulation of the isolated NBD dimer MJ0796 in different nucleotide bound states (ATP-bound, ADP in one binding site and ATP in the other, and the nucleotide-free state). From these simulations, the authors observed that the presence of ADP in one of the active sites promotes an enhanced (with higher number of H-bonds) NBD-NBD and NBD-ATP interactions in the opposite site, demonstrating that there is close correlation between binding sites in this dimer. Later in 2007, Jones and George [87] used theoretical methods (molecular dynamics simulations at 358K/50 atm and elastic network analysis) to study the isolated NBD MJ0796 monomer and dimer in different bound forms (ATP-bound, ADP-bound and nucleotide free state for the monomer, and the ATP and

ADP/ATP-bound form for the dimer). The dimer simulations revealed a large asymmetrical transition of one helical sub-domain, resulting in an asymmetric conformation of the dimer.

Molecular modeling studies have also been able to give us important insights into the dynamics of nucleotide driven NBD closure [94]. In 2006, Oloo and co-workers [94] used molecular dynamics simulations to study the effect of the presence of ATP in the binding sites of the maltose importer NBDs (MalK). In their simulations, the introduction of MgATP into the binding sites transformed a semi-open state into a closed state, whereas the open dimer evolved to a semi-open state. In contrast, in the absence of MgATP, the NBDs of both the open and semi-open states drifted further apart.

Recently, molecular dynamics techniques were used in order to understand the mechanisms of dimer opening upon ATP hydrolysis in the NBD dimer [96, 97]. In 2008, Wen and Tajkhorshid [96] applied molecular dynamics simulations to study the detailed mechanism of dimer opening in the MalK dimer. In this work, 50 ns MD simulations of four possible ATP and ADP plus Inorganic Phosphate combinations were performed. After 30 ns (in the fastest cases), it was possible to observe opening of one or both binding sites (for all systems with the exception of the one with two ATPs) and to identify the atomic details associated to this process. Later in 2009, Jones and George [97], also used molecular dynamics simulations to study the opening of the isolated NBD MJ0796 dimer. After extending the dimer simulations performed previously [87], the authors were able to observe the rotation of the helical sub-domain relative to the catalytic domain, which allowed the opening of the ADP-bound binding site.

Additionally, molecular dynamics simulations were used to study the mechanism by which NBDs power transport in the ferric iron uptake transporter (FbpC) from the *Neisseria gonorrhoeae* [56]. From the simulations, Newstead and co-workers [56],

claimed that the NBD dimer FbpC closed structure has higher energy than their respective open state, which can have significant implications in the mechanism of power generation in ABC transporters.

1.4.1.2. Full-length ABC transporters

Despite the rapid advances in computational power, molecular dynamics simulations of complete ABC transporters still present unique obstacles, mainly due to the large systems size. These gigantic systems require enormous amounts of computer processing and data storage. Nevertheless, over the last years, several works aiming to study the dynamics of complete ABC transporters have been reported [93, 95, 162-170].

The first molecular dynamics simulation study of a complete ABC exporter was reported in 2003 and in this work, Campbell and co-workers [162] studied the stability of the MsbA monomer and dimer embedded in a membrane mimetic environment (octane slab). In 1 ns of unrestrained simulation, the dimeric system showed considerable loss of helical secondary structure and underwent significant conformational drift, calling in to question the correctness of the crystallographic structure. Later, in 2006, Haubertin and co-workers [164] simulated for 50 ns (16 ns restrained and 34 ns unrestrained) the reconstructed MsbA TMD dimer inserted in a palmitoyloleoylphosphatidylethanolamine (POPE) bilayer. This simulation provided valuable insights into the structural relationships between the protein and its surrounding lipids. Additionally, in 2008, Ivetac and Sansom [166], used molecular dynamics simulations in order to evaluate the quality of a membrane protein structure or model. In this work, a 20 ns simulation of the MsbA model embedded in a lipid bilayer was compared to a 20 ns simulation of the bacterial Vitamin B_{12} importer (BtuCD). The BtuCD importer demonstrated good conformational stability during the simulation, whereas the MsbA exporter showed dramatic structural deformations. From the results, the authors propose that MD studies can be used

as a quality control tool for membrane protein structural biology, mainly as it provides a way to explore the global conformational stability of the protein. Recently, in 2010, Weng and co-workers [170], used targeted molecular dynamics simulation methods to determine the atomic details of the conformational transition from the outward-facing to the inward-facing states in the bacterial MsbA exporter. The molecular dynamics trajectories suggested that the disruption of NBD dimer interface is the first event that initiates the conversion of an outward- to an inward-facing state. The NBD dissociation induces, through the x-loops, rearrangements in the transmembrane helices at the cytoplasmic side and the periplasmic side sequentially.

Until now, five independent MD studies were performed in the Vitamin B_{12} importer (BtuCD) from *E. coli* [93, 95, 163, 165, 169]. In the first work, Oloo and Tieleman [93] simulated the BtuCD importer in two distinct nucleotide bound states (nucleotide free- and ATP-bound) embedded in an explicit lipid bilayer for 15 ns. From the nucleotide bound simulation, the authors demonstrate that the presence of ATP in the binding sites progressively draws the two NBDs towards each other. The NBDs movement induces conformational rearrangements in the TMDs, in particular, in the periplasmic end of the transporter. However, these simulations did not include the periplasmic substrate binding protein (BtuF) or the allocrite (vitamin B_{12}), which are thought to play an important role in the transport mechanism. In the second work, Tanizaki and Feig [163], simulated the BtuCD importer inserted into an implicit membrane environment. In this study, the dielectric constant of the membrane varied as a function of the distance from the center of the bilayer. During the 5 ns of the simulation, BtuCD remained structurally stable, but there were some indications (lower B-factors) of slightly suppressed protein mobility using this implicit membrane model. In the third work, Sonne and co-workers [165], studied nucleotide binding and release in the BtuCD importer, using perturbed elastic network calculations and biased molecular dynamics simulations. Both methodologies predicted that nucleotide release opens the cytoplasmic gate

whereas nucleotide binding has the opposite effect. However, the rearrangements in the cytoplasmic gate region observed were not large enough to allow vitamin B_{12} diffusion from the transmembrane pore into the cytoplasm. In the fourth work, Ivetac and co-workers [95], used multiple molecular dynamics simulations to explore the conformational dynamics of the various components of the BtuCDF complex, in order to understand the changes induce by nucleotide binding and the role of the BtuF in the complex. From the comparison of the seven simulations performed, the authors observed, that in the isolated NBDs, ATP binding induces the closure of the NBD dimer in a symmetrical fashion, but not in the complete transporter. In the BtuCD system, the NBDs closure only occurred after BtuF binding to the complex. In the fifth work, Kandt and Tieleman [169], used a combined protein–protein docking and molecular dynamics simulation approach to study the substrate sensitivity in the BtuCDF complex. Two possible holo-BtuCDF docking complexes were built, differing by a 180º rotation of the SBP. Both complexes were embedded in a lipid environment in order to investigate their dynamics and the transporter conformational response to the presence and absence of the BtuF, vitamin B_{12}, and MgATP. From the 28 independent MD simulations performed, it was possible to conclude that the holo-BtuF stabilizes the open BtuCD conformation and the BtuF or vitamin B_{12} removal induces BtuCD closure.

Recently, Weng and co-workers [167] reported an elastic normal-mode analysis of the metal-chelate-uptake ABC transporter HI1470/1. Using this method, it was possible to gain new insights into the intrinsic asymmetric conformational flexibility of the transporter and to demonstrate that the inward-outward transition occurs through the asymmetric motion of the transporter individual subunits.

Lastly, in 2010, Aittoniemi and co-workers [168], used conventional molecular dynamics simulation methods to study the transient asymmetric transitions in the nucleotide binding sites in a pre-hydrolytic state of the homodimeric multidrug exporter Sav1866. From their multiple MD simulations, it was possible to observe

that the initial symmetrical MgATP-bound state rapidly (~10 ns) evolves to an asymmetric NBD conformation and the hydrolytically favorable conformation was only detected in one of the nucleotide binding sites.

1.4.2. Structure prediction

Homology modeling is a computational method that allows the construction of an atomic resolution tridimensional model for a protein (target) based on its amino-acid sequence and in the experimental three-dimensional structure of a related protein (template). However, this tool is highly dependent on the sequence identity between the target and the template. For the NBDs, which share a high sequence identity between ABC transporters, the use of comparative modeling techniques is completely feasible and acceptable. In contrast, the sequence identity between the transmembrane domains is very variable and can range from a reasonable value, to little or no sequence identity at all. Since the accuracy of the homology model is highly dependent on the sequence identity between target and templates, a healthy degree of skepticism should be used when analyzing the constructed models. Over the last years and despite the limitations of the comparative modeling techniques, numerous homology models were published for several complete ABC transporters (e.g [175-179]).

1.5. The ABC transporters studied

As previously stated, the class of ABC transporters is one of the largest classes of known transporters [5, 180], and the members of this family are present in all kingdoms of life. Given the conserved ABC architecture (two TMDs plus two NBDs) across the entire ABC superfamily, it is reasonable to expect a common hydrolysis and translocation mechanism for ABC transporters in general (independently of the transport direction). With the objective of testing such hypothesis, we decided to study some examples of importers and exporters. The model ABC importer used was the maltose transporter ($MalFGK_2E$) and the model exporters chosen were the Sav1866 and the MJ0796. Throughout this section, we will describe in more detail, both the structural and the biochemical features of our studied proteins.

1.5.1. ABC Importers: $MalFGK_2E$

One of the best functionally characterized ABC transporter is the maltose uptake system from *E. coli.* This import system translocates maltooligosaccharides, up to seven glucose units [181], and it is composed of a periplasmic maltose binding protein (MBP or MalE), two different integral membrane proteins (MalF and MalG) and two copies of an ATPase protein (MalK) [182]. A large amount of experimental information (such as mutational and X-ray studies) is available for this transporter, giving us relevant insights in its structure and functioning. Moreover, there are several high resolution crystal structures available for the MalK domains [52, 53] and two structures for the complete importer system [62, 142].

For the MalK dimer, structures for three different nucleotide-bound states have already been reported [52, 53] (see Figure 9): the nucleotide-free, the ATP- and the ADP-bound states. In the nucleotide-free state, which is considered the resting state, the NBDs are well separated and the dimer is exclusively maintained through

contacts in the regulatory domains. In the ATP-bound state (pre-hydrolysis state), the NBDs dimer is closed, with the two nucleotide molecules bound at the interface of the monomers. In the ADP-bound state (post-hydrolysis state), the two NBDs are again separated, similarly to the resting state. Additionally, the $MalFGK_2$ importer was also crystallized in two clearly distinct states of the transport cycle: an inward- (Figure 13A) and an outward facing (Figure 13B) conformation.

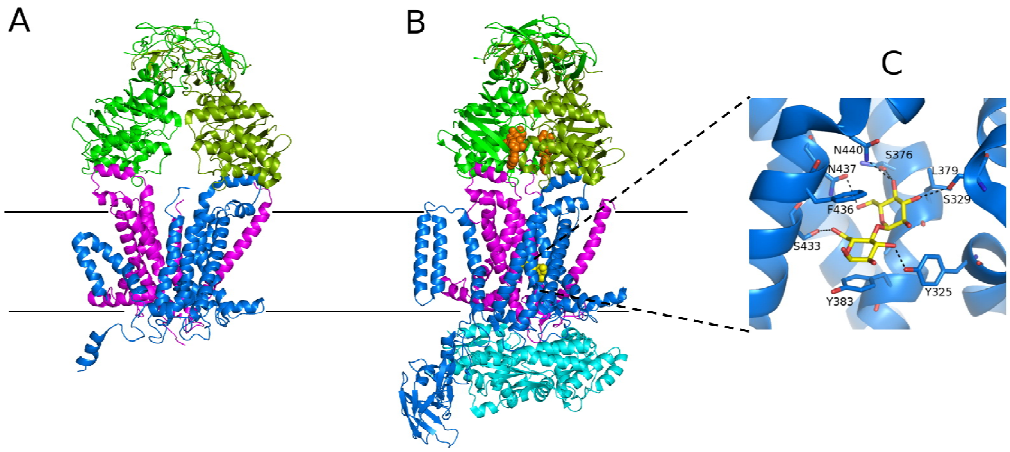


Figure 13- X-ray structures available for the complete maltose importer system ($MalFGK_2$) from *E. coli* [142; 162]. The horizontal black lines depict the probable location of the membrane based on the location of hydrophobic residues. Cartoon diagram for the **A**- inward-facing conformation (resting state) [142], and **B**- for the outward-facing catalytic intermediate conformation bound to its MBP (MalE) [183]. In these images, the ATPase domains are colored in green, the transmembrane domains in blue (MalF) and magenta (MalG), and the MBP (MalE) in cyan. The ATP nucleotides and the maltose allocrite, when present, are highlighted in orange and yellow, respectively. **C**- A detailed view of the substrate-binding site with maltose bound. The hydrogen bonds are indicated by dashed lines.

In the outward-facing conformation (Figure 13B), the MalK dimer is in a closed form, with two ATP molecules bound along the interface, in contact with residues from the Walker A/B motifs of one monomer and the LSGGQ motif of the opposite monomer. The NBDs present a structure essentially identical to that of the isolated

ATP-bound MalK dimer [52]. The structures of the two TMDs, MalF and MalG, are formed by eight and six transmembrane helices, respectively, and, surprisingly, the MalG C-terminal region is inserted into the MalK dimer. Part of MalF is located in the periplasmic region and extends about 30 Å away from the membrane surface. Additionally, in this structure, the maltose-binding protein (MBP or MalE) is also bound (in an open conformation) to the periplasmic side of the transporter, contributing to the formation of a large occluded cavity, located at the interface of the TMDs. The MalE C-lobe interacts mainly with MalF, whereas the MalE N-lobe interacts with MalG. In this case, the translocation pathway is completely shielded from the membrane by the MalF and MalG transmembrane helices, and from the periplasm by the open apo MalE. The maltose substrate is also located at the base of the transmembrane cavity (approximately halfway across the membrane) and it is associated, by ring stacking and hydrogen bonding interactions, exclusively to MalF (see Figure 13C). The NBD-TMD interface is primarily formed by contacts between a short TMD cytoplasmic helix, named the “coupling helix’ [70] (lying approximately parallel to the membrane), and a surface cleft formed between the MalK RecA-like and helical sub-domains.

In the inward-facing conformation (Figure 13A) [142], and contrary to the outward-facing conformation, the MalK dimer is in the resting state, with the nucleotide-binding interface opened and no MalE present. In this case, the MalF and MalG domains are separated in the membrane inner leaflet (with an aqueous channel between them) and packed together in the outer leaflet, shielding the maltose-binding site from the periplasm.

A comparison between the two distinct transport conformations structures described above shows that all domain interactions have changed, with the exception of the MalK regulatory domains. In the MalK-MalG/MalF interface, the rotation of the helical sub-domain region allows the movement of the TMDs relative to NBDs, without disrupting the interface between them. Moreover, the TMDs

conformational changes between the two states involve significant rigid-body rearrangements of the core transmembrane helices, leading to an alternating access to the cytoplasm [142].

1.5.2. ABC Exporters: Sav1866 and MJ0796

During this thesis two model exporters were studied in detail: the Sav1866 and the MJ0796 exporter. In the second case, given that a complete transporter X-ray structure does not exist, only the ATPase subunits were simulated.

Sav1866 exporter

The Sav1866 is a bacterial exporter from *Staphylococcus aureus* and it is an homologue of the multidrug exporters [58], such as MDR1 and TAP1/TAP2. Additionally, the Sav1866 ATPase activity is stimulated by some of the human MDR1 [184] and *Lactobacillus lactis* LmrA [154] substrates, such as doxorubicin, vinblastine or the fluorescent dye Hoechst 33342.

Until this moment, two crystal structures for the Sav1866 exporter are available (one with ATP- and the other with ADP-bound) [58, 59], both in the outward-facing conformation (see Figure 14). An explanation for the similarities between the ATP and the ADP-bound conformations is that the purification and crystallization conditions, in particular the presence of detergent, shifted the Sav1866 conformational equilibrium towards the ATP-bound state. The ATP-bound structure of the multidrug exporter Sav1866 was the first high-resolution structure reported for an ABC exporter [58] and it revealed a remarkably intricate fold and topology (see Figure 14). In particular, the TMDs are not simply aligned side by side but rather they are embracing each other in a complex way.

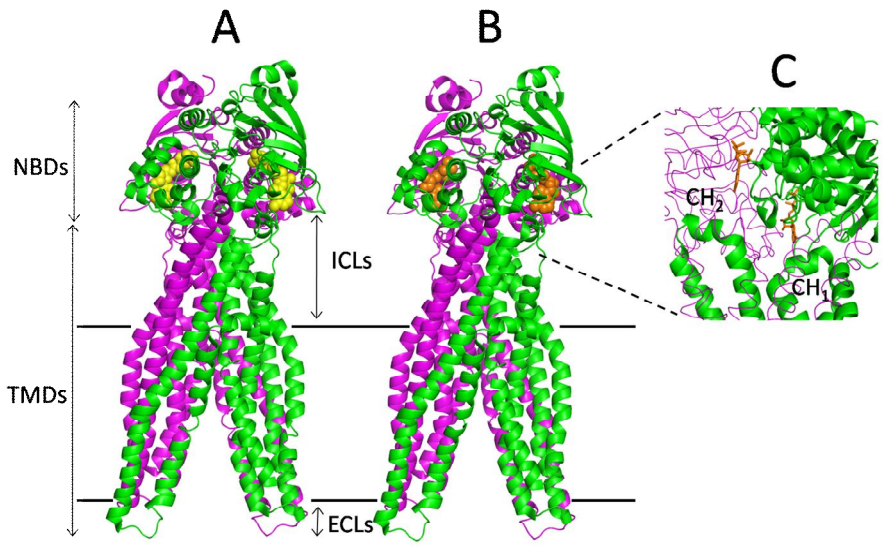


Figure 14- X-ray structures available for the Sav1866 homodimer [58, 59]. Two Sav1866 crystal structure are represented in this image: (**A**) the AMP-PNP bound [59] and (**B**) the ADP-bound state [58]. The two chains are colored in green and magenta while ATP and ADP are shown as yellow and orange spheres, respectively. The horizontal black lines depict the probable location of the membrane based on the location of hydrophobic residues. **C**-close-up view of the Sav1866 inter-domain interface (formed essentially by two coupling helices, CH_1 and CH_2, per monomer).

In exporters, the TMDs are generally fused to the NBDs, generally with the TMDs preceding the NBDs, forming a “half” transporter. Sav1866 is a homodimer of “half” transporters, and each subunit contains an N-terminal TMD, with six helices, and a C-terminal NBD, with the canonical fold of the ATPase domains. In both Sav1866 crystal structures, the NBDs are in closed conformation, with the nucleotides (ATP in [59] and ADP in [58]) bound at the interface, and the TMDs in an outward-facing conformation. The TMDs consist of 12 transmembrane helices arranged in two distinct “wings”. Each wing is formed by the first two helices (helix 1-2) from one TMD and the remaining four (helix 3-6) from the other TMD. The transmembrane segments are connected to each other by long intracellular loops (ICLs) that extend

the helical secondary structure beyond the membrane, and protrude ~25 Å into the cytoplasm. In the Sav1866 structures [58, 59], the inter-domain transmission interface is formed by two intracellular coupling helices, one (CH_1) contacting the NBDs of both subunits, and the other (CH_2) interacting solely with the NBD of the opposite subunit. This interface is different from the one identified for importers, where the coupling helices contact a single NBD.

The Sav1866 central cavity (located between the TMDs) is relatively hydrophilic, which was interpreted as an extrusion pocket, with little or no affinity for the allocrites (usually hydrophobic) [14].

Over the last years, the Sav1866 exporter, which is thought to represent the core architecture of ABC exporters, has been widely used as a model and a template for the comparative modeling of relevant human ABC members (e.g. [149, 185-187]).

MJ0796 exporter

The second ABC exporter studied in this thesis was the MJ0796 transporter. In this case, and contrary to the Sav1866 exporter, only the ATPase domains were studied in detail, mainly due to the fact that no crystal structure is available for the complete transporter.

The MJ0796 is a member of the o228/LolD [188] transporter family found in the hyperthermophilic Archaeon *Methanococcus jannaschii* [80, 188, 189]. Two X-ray structures for the ATPase domains are available: an ADP-bound monomeric [36] and an ATP-bound dimeric [51] state. The ADP-bound monomer was reported in 2001 by Yuan and co-workers [36]. In this structure, the monomer presents a typical bilobal configuration (previously described). The ATP-bound dimer was reported in 2002 by Smith and co-workers [51] and it convincingly showed to be the physiologically correct interface for the NBD monomers. Upon the comparison of the ATP- (dimer) and ADP-bound (isolated monomer) MJ076 NBDs, the main

difference observed is the ABC signature motif movement (~7Å), based on a 17 degree rigid-body rotation of the α-helical sub-domain (see Figure 15).

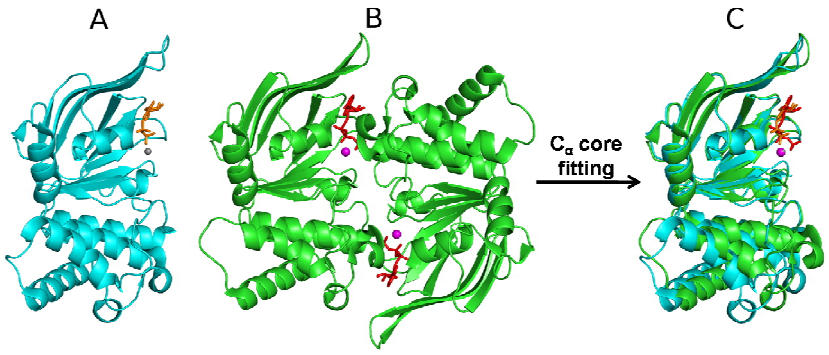


Figure 15- Conformational differences between the ADP- (pdb code: 1F3O) [36] and the ATP- bound (pdb code:1L2T) [51] MJ0796 monomers. **A**-Monomeric MgADP-bound wild type structure [36]. The ADP molecule is represented in orange sticks while the Mg is a gray sphere. **B**- Dimeric NaATP-bound E171Q mutant structure [51]. ATP is depicted as red sticks, while Na is represented as a magenta sphere. **C**- Superimposition of the ADP-bound wild type structure [36] to the E171Q mutant ATP-bound structure [51]**,** after least-squares fitting to the core domains (comprising the residues forming β-sheet$_{1\text{-}4}$, β-sheet$_{6\text{-}9}$ and α-helix$_1$)

2. Theory and Methods

2.1. Introduction

"Molecular modeling" is a term frequently used to identify a group of theoretical and/or computational models that provide insights into the behavior of a molecular system [190]. These models normally use a simplified description of the system based on mathematical and physical methods [190]. There are several different models, each providing different levels of approximation to the physical properties of the molecular system. These models differ in the level of modeling, that is, in the degrees of freedom they can tackle (nuclei and electrons or atoms), the properties/processes they can reproduce and the time scale allowed [191]. Two of the most used models in the biomolecular modeling field are shown in Table 2.

Table 2-Two of the most used models in computational biochemistry and molecular biology. This table was adapted from [191].

Model	Degrees of Freedom	Properties, Processes	Time scale
quantum mechanics (ab initio, density functional, semiemperical and valence bond methods)	Nuclei and electrons	Ground and excited states, reaction mechanisms	no time scale, picoseconds
classical mechanics (Molecular Dynamics, Monte Carlo)	Atoms	System properties from ensemble averages	nanoseconds

An example of a very complete and accurate model is quantum mechanics, which explicitly treats nuclei and electrons and allow the study of reaction mechanisms. However, many of the questions we would like to answer cannot be treated using this method due to large number of particles that need to be treated, which makes it difficult to reach sufficiently large timescales. In order to overcome these limitations due to system size and limited timescales, simplified methods named molecular mechanics/dynamics methods were developed, which use physical simplifications to treat the conformational behavior of molecules. The results obtained with these simplified models can be then compared (whether in a quantitative or qualitative way) with known experimentally determined properties, which then validate, or not, the model used. After model validation, this can be use as a predictive tool for unknown or unmeasurable quantities or phenomena [192].

In this section, we will give a brief overview of the principles underlying the theoretical methodologies used in the ABC transporter studies reported throughout this thesis. A more detailed description of each one of these techniques can be found in several reference books [190, 193, 194].

It is important to refer that in the molecular mechanics/dynamics methods used, the nuclear motion of the particles obeys the laws of classical mechanics [194], the electrons are assumed to be in the ground state [194] and they are not explicitly treated (their effects are incorporated in the nuclei) [190] and the energy of a system is calculated based on the nuclear positions only [190]. The properties obtained from these simplified molecular models can then be related to the properties measured experimentally with the support of the theoretical principles of statistical mechanics [190]. However, it is important to emphasize that molecular mechanics/dynamics cannot provide any information of a property which depends directly upon changes on the electronic distribution in a molecule (for example: when chemical reactions take place changing the covalent structure) [190]. For

these specific cases, the reactive part of the system needs to be treated explicitly by quantum chemical methods [194].

2.2. Molecular Mechanics/Dynamics

2.2.1. Empirical Force Field

A simple mechanical description of the atomic interactions in a molecular system can be done by a potential energy function that describes the bonded and non-bonded interactions between the atoms. These energy functions are known as force fields (FF). A FF, generally, consists of several potential energy terms representing the covalent interactions between atoms (such as angle bending, bond stretching and angle torsion) and the non-bonded interactions between atoms in different molecules or between atoms in the same molecule separated for more than two covalent bonds [191]. The FF is often defined as the sum of these terms (equation 2.1) after calibration [191]:

$$V(\boldsymbol{r}^n) = V_{\text{bonds}} + V_{\text{angles}} + V_{\substack{\text{improper}\\\text{dihedrals}}} + V_{\substack{\text{proper}\\\text{dihedrals}}} + V_{\substack{\text{van.}\\\text{der}\\\text{Waals}}} + V_{\text{electrostatics}} \qquad (2.1)$$

Over the last decades, several FF have been developed based on different principles, and capable of reproducing different properties of a molecular system. “All-atom” FFs are available (such as the OPLS-AA FF [195]), where all atoms in the system (including hydrogens) are explicitly treated. Additionally, there are the “united-atom” FFs (such as the GROMOS96 43A1 FF [196, 197]), where the aliphatic hydrogens are incorporated into the atom to which they are bond [194]. Lastly, there are also the “coarse-grain” FFs where the functional groups in the molecular systems are represented by simpler coarse particles rather than individual atoms (such as the MARTINI FF [198]). The choice of the FF is not a straightforward task and depends directly on several factors, such as, the simulation time and the properties we want to study.

For the work reported in this thesis, we have used both the GROMOS96 43A1 [196, 199] (for the study of the soluble ATPase domains) and the GROMOS96 53A6 [197, 200] (for the study of the membrane-protein complexes) force fields. Both FFs use the "united-atom" approach described above, and due to the reduction in the number of interactions sites, is able to decrease the computational time. In the GROMOS FFs, all the polar hydrogens (bound to nitrogen, oxygen or sulphur) are explicitly treated as well as some aromatic hydrogens. In these FFs, the potential energy of a system $V(\boldsymbol{r}^n)$ composed of n particles is determined as a function of the atomic position ($\boldsymbol{r}$) of each particle [191]:

$$V_{\text{bonds}} = \sum_{\text{bonds}} \frac{1}{4} k_{\text{b}} (b^2 - b_0^2)^2 \tag{2.2}$$

$$V_{\substack{\text{angles} \\ \text{bond} \\ \text{angles}}} = \sum \frac{1}{2} k_\theta (\cos\theta - \cos\theta_0)^2 \tag{2.3}$$

$$V_{\substack{\text{improper} \\ \text{dihedrals}}} = \sum_{\substack{\text{improper} \\ \text{dihedrals}}} \frac{1}{2} k_\xi (\xi - \xi_0)^2 \tag{2.4}$$

$$V_{\substack{\text{proper} \\ \text{dihedrals}}} = \sum_{\substack{\text{proper} \\ \text{dihedrals}}} k_\varphi \left[(1 + \cos(\delta)\cos(m\varphi)\right] \tag{2.5}$$

$$V_{\substack{\text{van} \\ \text{der} \\ \text{Waals}}} = \sum_{\substack{pairs \\ i,j}} 4\varepsilon_{i,j} \left[\left(\frac{\sigma_{i,j}}{r_{i,j}} \right)^{12} - \left(\frac{\sigma_{i,j}}{r_{i,j}} \right)^{6} \right] \tag{2.6}$$

$$V_{\text{electrostatics}} = \sum_{\substack{pairs \\ i,j}} \frac{q_i q_j}{4\pi\varepsilon_0 \varepsilon_r} \frac{1}{r_{i,j}} \tag{2.7}$$

As can be seen, the potential energy function is formed by six individual terms that describe the bonded (equations 2.2 to 2.5) and non-bonded (equations 2.6 and 2.7) contributions.

A schematic representation of both the bonded and non-bonded contributions is depicted in Figure 16.

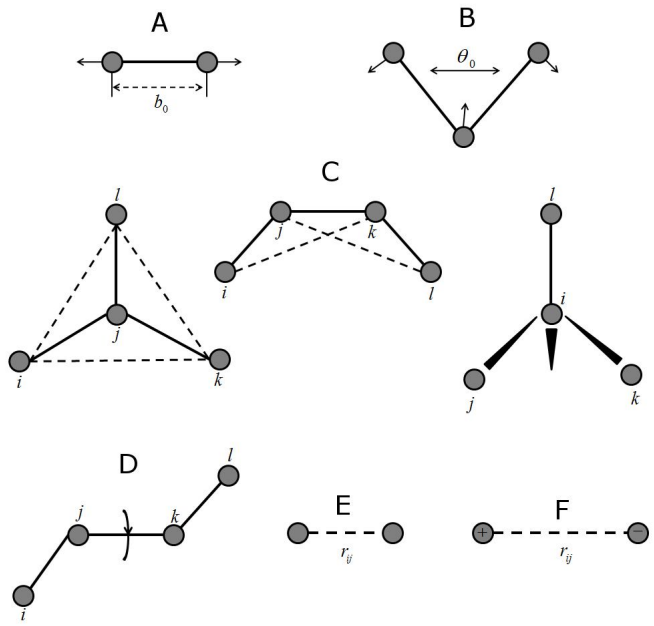


Figure 16- Schematic representation of both the bonded (A-D) and non-bonded (E-F) contributions to a molecular mechanics force field: **a**) bond stretching (equation 2.2); **b**) angle bending (equation 2.3); **c**) several types of improper dihedrals (equation 2.4); **d**) proper dihedral (equation 2.5), **e**) van der Waals interactions (equation 2.6) and **f**) electrostatic interactions (equation 2.7).

2.2.2. Bonded interactions

The covalent bond stretching contribution (Figure 16A) is described by equation 2.2. In this case, the energy between two bonded atoms is defined by a force constant (k_b) and by the difference between the bond length values (b) and a

reference bond length (b_0). The bond stretching is modeled by a harmonic quadratic potential that does not allow bond breaking.

The equation 2.3 describes the bond angle bending (three-body) potential (Figure 16B) in the system. In this case, the angles between atoms are also modeled by a harmonic potential where k_θ is the force constant and θ_0 the reference angle.

Equation 2.4 describes the improper dihedral angle (four-body) potential (Figure 16C) that are not allowed to make transitions. This potential maintains the atoms restricted to a plane and avoid transitions in to a configuration of opposite chirality. These interactions are modeled by a harmonic potential where k_ξ is the force constant and ξ_0 the reference angle.

Finally, equation 2.5 represents the proper dihedrals angle potential (Figure 16D), and describe a 360 degrees rotation of a central bond in a four component system. In this case, the proper dihedrals are modeled by a sinusoidal term where k_φ is the force constant, m the multiplicity and φ the phase shift. The multiplicity describes the periodicity of the trigonometric function, whereas the phase shift describes which are the dihedrals with maxima energies.

2.2.3. Non-bonded interactions

The last two terms in equation 2.1 represents the non-bonded contributions (fFigure 16E and Figure 16F) made by the van der Waals (equation 2.6) and the electrostatic Coulombic (equation 2.7) interactions. These interactions are calculated for all atom pairs separated by at least two or three bonds [197], and they are modeled as a function of the distance between atoms. The van der Waals interactions are represented by the Lennard-Jones (or 12-6) function described in

equation 2.6, which combines a repulsive and an attractive term (Figure 17). As can be seen, when two atoms are too close to each other the energy rapidly increases due to the overlapping of their electronic clouds. This region of the graphic corresponds to the repulsive (or C_{12}) potential. As the distance between the atoms is increased, the energy decreases until it reaches a minimum [190]. From this point forward the attractive potential starts to be more effective over the repulsive one. When the two atoms are too far apart the interaction energy is negligible [190].

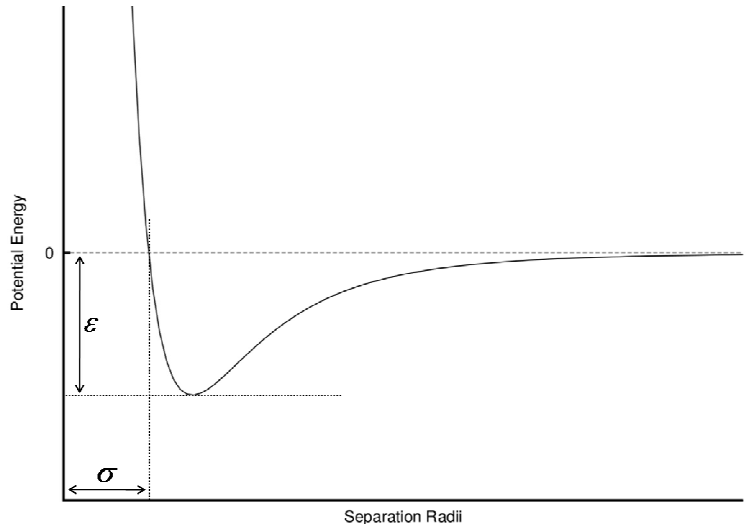


Figure 17- Representation of the Lennard-Jones potential. Repulsive (C_{12}) and attractive (C_6) components.

In Equation 2.6, σ corresponds to the distance between the two atoms for which the potential interaction energy is zero, and ε is the minimum potential energy for a pair of atoms. It is important to refer that both σ and ε are unique for each pair of atoms. Typically, a FF contains ε and σ values for a pair of equal atoms. However, for a system containing N different type of atoms, $N(N-1)/2$ sets of parameters of the interaction of unlike atoms would be required. So, the

determination of σ or ε can be difficult and time consuming, and in order to overcome this limitation it is assumed that these parameters for different pair of atoms types can be obtained from parameters of equal atoms using mixing rules [190] (such as Lorentz-Berthelot rules or simple geometric averages).The van der Waals interactions are presumably calculated for all atom pairs, but, in order to reduced the computational cost of the simulation, the use of a cutoff radius is common. In this case, a list of neighbor atoms lying inside the cutoff is created (and then updated every so many simulation steps) and all the atom pairs separated by a distance larger than the cutoff distance will be ignored and have a van der Waals interaction energy equal to zero. In our simulations, a twin-range cutoff scheme (see Figure 18) was used for the Lennard-Jones interactions (as well as for the electrostatic Coulombic interactions).

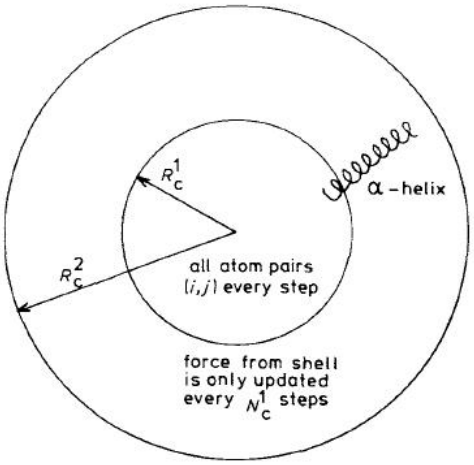


Figure 18- Twin range method. This figure was adpated from [192]

In this method, two cut-off radii are specified, R_C^1 and R_C^2. All atom pairs below the lower cutoff are calculated every step, whereas interactions due to atoms located between R_C^1 and R_C^2 are evaluated only when the neighbors list is updated (every

N_C^1 time steps) and are kept constant between these updates. In our simulations with the GROMOS FF [196, 197, 199], R_C^1 and R_C^2 was 0.8 and 1.4 nm, respectively.

The last component in equation 2.1 is the electrostatic interaction calculated using Coulomb's law. In equation 2.7, the Coulombic interactions between two atoms (i and j) depend on the atomic partial charges of the interacting atoms (q_i and q_j), the distance between them ($r_{i,j}$) and on the vacuum (ε_0) and relative permittivity of the medium (ε_r). Unlike the previously described van der Waals term, these interactions do not rapidly decay with the distance (they decay proportionally to $1/r_{i,j}$). However, in order to reduce the computational cost of the columbic interactions, and similarly to the procedure for the van der Waals term, the use of a cutoff criterion is frequent. Like described for the van der Waals term, when a cutoff radius is applied, a list of neighbor atoms lying inside the cutoff is created and all outside-cutoff remaining interactions are ignored. However, this simple and direct truncation can introduce large energy discontinuity in the systems, which can lead to serious artifacts and to instability in the simulations [201]. Nowadays there are available other more sophisticated long-range electrostatic treatments, which better and more accurately describe the electrostatic interactions in molecular systems, such as Ewald summation [202] (or some variants like PME [203] methods) or reaction-field methods [204]. So, the long-range contributions to the electrostatic energy can be evaluated by two methods: the lattice-sum methods or the continuum methods. Examples of lattice-sum methods are the Ewald summation [202] and all its fast Fourier variants (such as particle-mesh Ewald [202, 203, 205]). In the lattice-sum methods, the system is considered to be positioned in a central box, surrounded by an infinite number of copies of itself and, by this reason, an artificial periodicity is introduced in the system. These methods are considered the most adequate to accurately include the effects of long-range forces in a simulation.

However, they have some known weakness, such as, the enhance of artifacts arising from periodicity [190] and they are computationally very expensive [190]. An alternative to lattice-sum methods, are the continuous corrections, such as the Reaction-Field method [204]. In this type of approximations the system is divided in two distinct parts: the first one corresponds to the inner region (inside the cutoff region where the atomic charges are explicitly treated) and a second one corresponds to an outer region (where the medium beyond the cutoff radii is modeled by a homogeneous continuous medium with a dielectric constant, ε_{rf} , and ionic strength, I). The electrostatic interaction energy, in this case, is determined by the sum of these two values. The main weakness of this method arises from the fact that outside-cutoff interactions are modeled as a homogeneous solvent medium of dielectric constant ε_{rf} (even if they are not only solvent) [206].

Throughout this thesis, both the particle-mesh-Ewald (PME) method and the reaction-field method will be used. When using the reaction-field method for the calculation of the electrostatic Coulombic interactions, a twin-range cutoff scheme, similar to the one describe previously for the van der Waals interactions, was used.

2.2.4. Solvation and Boundary Conditions

Another important aspect when performing protein simulations is the correct characterization of the environment where the proteins are inserted. Most proteins are simply simulated in an aqueous medium, but others also need to be inserted into a lipidic environment, additionally to the aqueous solvent. Most simulations nowadays treat the solvent explicitly; although the evaluation of the solvent-solvent interactions is very time consuming (the largest part of the simulation time is spent evaluating these interactions). Alternative solvent models are available in which the solvent is implicit [207], such as the Generalized Born model [208]. However, all works reported in this thesis used explicit solvent.

A typical biomolecular system is formed by 10^5 to 10^6 atoms, which is a very small size system (when compared to the Avogadro constant). Due to the system small finite size, the way system boundaries are treated is essential to a simulation, since it enables the determination of macroscopic properties from a relatively small number of particles. The classical and more commonly used way to minimize the edge effects in a finite system is to use periodic boundary conditions (PBC). In this algorithm, the system is considered to be replicated in all three dimensions, which means that the system becomes surrounded by several copies of itself. This algorithm enables the simulation to be performed using a central box only, in such a way that the forces applied to the particles are similar as if they were in a bulk. The use of PBC, although minimizing the boundary effects, introduces artificial periodicity (meaning that in fact a crystal is being simulated) and removes the conservation of the angular momentum of the system.

2.2.5. FF parameterization

As previously described, a FF consists of potential-energy terms (equation 2.1) representing the bonded (bond-streching, angle bending, proper an improper dihedral torsion) and the non-bonded (van der Waals and Coulombic) interactions between atoms in the same molecule or in different molecules. Most biomolecular FFs possess hundreds of parameters, taken in most cases from experimental data of small molecules and from theoretical calculations, which can model the structural and thermodynamic properties of a molecular system. In the GROMOS FF, the parameters for the reference bond lengths and angles are obtained from crystal structures of small molecules whereas, the corresponding vibrational force constants are taken from infrared spectroscopic data of small molecules [191]. Additionally, the parameters for both the partial atomic charges and the proper dihedral angles are obtained from quantum-chemical calculations [191]. The non-bonded interaction parameters, C_6 and C_{12} are determined by fitting heats of

vaporization, densities of pure liquids and free energies of solvation of small molecules in the condensed phase [197, 200]. The parameterization of a FF is a formidable task, since someone has to find appropriate and consistent values for all the parameters in order to reproduce a specific set of properties (for an overview of the parameterization procedures see [209]). Calibration of a FF involves, sometimes, long simulations in order to generate good configurational ensembles.

From the discussion of the previous sections it is now clear that the “perfect” FF does not exist and the adequate FF choice depends on the type of molecular system and in the type of properties one is interested in.

2.2.6. Computer Simulation methods

Computer simulation methods are nowadays routinely applied in the biochemistry and molecular biology fields in order to understand and predict several properties of a biomolecular system. These methods allow the study of several systems and their properties throughout the use of techniques that consider small replicates of the macroscopic system with manageable numbers of atoms or molecules. A simulation, therefore, generates a group of representative configurations of these small replicates (ensemble of configurations) in a way that the adequate values of thermodynamic and structural properties can be obtained. Moreover, some simulation techniques (such as for example, the Molecular Dynamics simulations [190]) can also enable the determination of the time-dependent behavior of a system, allowing a detailed picture of how the system changes from one conformation to another.

Various methods are available for searching the configuration space and for generating a conformational ensemble, each with their own particular weaknesses and strengths. Two of most common searching methods are the Molecular Dynamics [190] and the Monte Carlo methods [190].

2.2.7. Molecular Dynamics simulation

The Molecular Dynamics (MD) method allows the generation of a conformational ensemble through the integration of Newton's equations of motion over time [190, 207]. This method involves the iterative numerical calculation of the instantaneous forces acting on a biomolecular system and the generation of a trajectory (configurations as a function of time) for all atoms in the system. The atomic positions for all atoms are obtained by solving the classical (Newtonian) equations of movement [190, 207]. This method is deterministic, which means that the state of the system at any future time can be predicted from its current state. The MD method allows us to sample and obtain, not only a representative ensemble of configurations for the system (if the simulation time is sufficiently long) but also the dynamical behavior of the system and the temporal relation between consecutives conformations.

Considering a system of n particles with mass m_i, the Newton's equations of motion states that there is a relationship between the atoms mass (m_i), the acceleration ($\boldsymbol{a}_i$) and an applied force ($\boldsymbol{F}_i$) according to equation 2.8:

$$\boldsymbol{F}_i = m_i \cdot \boldsymbol{a}_i \tag{2.8}$$

The Newton's equations of motion can also be written in terms of velocities ($\boldsymbol{v}_i$) and atomic positions ($\boldsymbol{r}_i$) with respect to time (t):

$$\boldsymbol{F}_i = m_i \cdot \frac{d\boldsymbol{v}_i}{dt} = m_i \cdot \frac{d^2\boldsymbol{r}_i}{dt^2} \tag{2.9}$$

For each configuration of the system, the forces ($\boldsymbol{F}_i$) acting on every atom i can be calculated from the negative gradient of the potential energy function previously described in section 2.2.1 with respect to the position of the atom ($\boldsymbol{r}_i$):

$$\boldsymbol{F}_i = -\frac{\partial V(\boldsymbol{r}^n)}{\partial \boldsymbol{r}_i} \tag{2.10}$$

In practice, in a MD algorithm, the force is calculated taking the negative gradients of all potential energy components of a FF (equation 2.2-2.7) [193] and then use afterwards in the integration of the Newton's equation of motion (equation 2.9) , where velocities and positions at time $t + \Delta t$ are obtained using numerical methods, from their initial values at instant t and their variation after Δt. There are numerous algorithms for numerical integration of the equations of motion, such as the Verlet [210], the leap-frog [190, 211] or the Runge-Kutta [212] algorithms. In our case, all MD simulations performed used the leap-frog algorithm, which is the one implemented in the GROMACS MD package [213, 214] (the MD simulation program used throughout this thesis). Moreover, it is considered to be one of the most simple, efficient and accurate algorithms for the molecular simulations of fluid-like systems [192]. In the leap-frog algorithm, the positions ($\boldsymbol{r}_i$) at time *t* and the velocities ($\boldsymbol{v}_i$) at time $t - \frac{\Delta t}{2}$, as well as the force ($\boldsymbol{F}_i$) at time t, are used for the update of the velocities (equation 2.11) and positions (equation 2.12):

$$\boldsymbol{v}_i(t + \frac{\Delta t}{2}) = \boldsymbol{v}_i(t - \frac{\Delta t}{2}) + \frac{\boldsymbol{F}_i(t)}{m_i}\Delta t \tag{2.11}$$

$$\boldsymbol{r}_i(t + \Delta t) = \boldsymbol{r}_i(t) + \boldsymbol{v}_i(t + \frac{\Delta t}{2})\Delta t \tag{2.12}$$

In this algorithm, a choice of an adequate time step (Δt) must be made, and during Δt the forces are assumed to be constant. If we choose a time step too large, this can lead to instabilities in the integration algorithm and to the failure of the energy and momentum conservation. On the other hand, an excessively small time step will lead to longer calculation times. Δt is directly dependent on the highest frequency motions occurring in the system and one way used to increase Δt is to use algorithms that constrain some degrees of freedom (such as bond lengths or possible bond angle vibrations) that contribute to the most high frequency motions. Some common constraint algorithms are SHAKE [215], LINCS [216] (used to reset bond lengths to its reference values) and SETTLE [217] (especially developed to constrain rigid water molecules).The use of a constrain algorithm helps to significantly save computing time and usually, for a simulation with constrained bonds, a time-step of 2 fentoseconds is employed [192].

2.2.8. Metropolis Monte Carlo simulations

The Metropolis Monte Carlo (MMC) method [218] is an alternative to molecular dynamics simulations, and this technique is also able to generate an ensemble of configurations from which structural and thermodynamic data can be extracted. MMC simulations are a stochastic approach to generate a set of representative configurations under given conditions. In this method and contrary to MD simulations, there is no information about the time evolution of structural events, and each configuration depends only of its predecessor. In its simplest form, in the MMC algorithm, a set of variables (like the atomic position or a dihedral rotation) are randomly changed followed by acceptance or rejection, of the newly generated configuration based on its energy. If the energy of the new configuration is lower than the energy of its predecessor configuration, then the new configuration is accepted. If the energy of the new configuration is higher than the energy of its

predecessor, the probability of acceptance $\rho(\boldsymbol{r}^n)$ of the new configuration ($\boldsymbol{r}^n$) is given by the Boltzmann factor (equation 2.13):

$$\rho(\boldsymbol{r}^n) = e^{\{-\Delta V(\boldsymbol{r}^n)/k_B T\}} \qquad (2.\ 13)$$

In equation 2.13, k_B *is* the Boltzmann constant, T the absolute temperature of the system and $\Delta V(\boldsymbol{r}^n)$ is the change in potential energy between the two configurations. A random number between 0 and 1 is then generated and compared to the probability of acceptance. The new configuration is only accepted if the random number generated is smaller than $\rho(\boldsymbol{r}^n)$. Upon acceptance, the new configuration is used as the starting point to the next random displacement. This process is repeated until sufficient sampling statistics for the current temperature T is achieved. The Metropolis criteria ensure that the lowest energy states are thus generated with higher probability than high energy configurations.

In this thesis, MMC simulations were used in the determination of the protonation states of aminoacid side chains and also in the determination of the adiabatic energy profile of doxorubicin along the transmembrane pathway for the Sav1866 exporter.

2.2.9. Membrane Protein simulations

Membrane proteins are of high biological and medical relevance since they are key players in crucial processes, such as transport or energy conversion in cells. They are, when compared to soluble proteins, much harder to crystallize, essentially due to their hydrophobic surfaces (which are stabilized by the membrane where they are inserted in). Usually, the isolated protein is rather unstable and easily denaturates when removed from the lipid bilayer, which complicates crystal

production. However, over the last decade, and due to significant improvements in the experimental area, the number of high-resolution membrane protein structures, deposited in the Protein Data Bank, has rapidly grown.

Nowadays, the increasing sophistication of FFs, modern software and increasing computer power, enabled computer simulations to become a standard technique to study, with atomic detail, the dynamic behavior of membrane proteins. However, many questions such as the adequate choice of the lipid molecules composing the simulated membrane or the timescale necessary to simulate the phenomena of interest should be carefully and realistically considered prior to a simulation.

A lipid membrane is a self assembled structure in an aqueous environment, formed due to the hydrophobic effect. The lipids can be divided in three main classes: phospholipids, glycolipids and sterols. These classes are very diverse, whether in structure or in biophysical properties, and the lipids structure mostly differs in the polar head group and in the acyl chain composition. The membrane environment (which serves as a “solvent” for the hydrophobic transmembrane spanning regions) is very important for the distribution, organization and function of membrane proteins, through electrostatic and steric interactions [219], and by this reason, the proper choice of the lipid environment in which proteins are inserted is of great relevance. Nevertheless, due to the large number of lipid species available in cells (for example, erythrocytes are estimated to have more than 200 different lipids), the choice of an adequate lipid is not a trivial task. Moreover, the membrane thickness is also affected by membrane constitution (the type of lipids and the presence of other molecules, such as cholesterol), and it usually varies between 5-10 nm.

Although accurate models of pure phospholipids are an essential component of membrane protein simulations, the parameterization of such molecules is especially difficult, due to its highly complex phase diagram. In physiologically relevant conditions, a lipid bilayer should have a liquid-crystalline organization [220].

However, a small change in temperature can induce phase transition. Over the last years, significant progresses have been reported for the improvement of the basic lipid models (whether saturated or polyunsaturated) and they can, nowadays, accurately reproduce several experimental properties, such as the translational diffusion coefficients, the area and volume per lipid, and the order parameters. Additionally, and although most membrane-protein simulations are performed in a pure, one-component neutral phospholipid bilayer, in biology, the charged lipids make up about 10-30% of a membrane [221]. So, over the last years, the lipid mixtures field has also reported relevant advances, although the time scale required for the equilibration of the mixtures (at least hundreds of nanoseconds) is still an obstacle.

The number of atoms in a protein-membrane system can easily reach 10^5–10^6, which poses a major computational challenge. In order to reduce the computation time and to allow effectively longer simulations, some membrane mimetics [222, 223] (such as slabs of hydrophobic molecules) and implicit solvent models [224, 225] are also frequently used. However, although these models give us useful results, the detailed picture of the atomic protein-lipid is completely lost. One major limitation in membrane-protein simulations, is related to the size of the simulated system and the artifacts it can cause. While in experimental works the bilayer (whether planar arrays or vesicles) are measured in a micrometer scale, in a MD simulation and due to computational limitations, a system is on the nanometer-scale. The small scale of the lipid patches used in simulation introduces some known artifacts, such as finite-size effects or the suppression of large-scale phenomena like undulations. These subjects have been widely debated in literature over the last years (see for example [226]) as well as the ways to circumvent such limitations (see for example [226]).

In this thesis, all the membranes used were pure, one-component dimysristoylphosphatidylcholine (DMPC) bilayers formed by 512 molecules in the

liquid-crystalline state. The DMPC was chosen because it was a model phospholipid (and did not need to be parametrized) and its properties could be easily validated by comparison with the experimental data available.

2.3. Continuum Electrostatic Methods

Electrostatics plays a key role in biological processes (such as protein-ligand interaction or guidance of substrates towards an enzyme active site) due to their strength and long-ranged nature. Additionally, the protonation state of a biomolecule is crucial to its function and stability, and its electrostatic properties are influenced by pH and the ionic conditions. The approaches used to model electrostatic interactions in biological systems can be divided in two groups: the first one that explicitly simulate all atoms in the system (and are the previously described MM/MD methods [190-192]) and a second one which treats the solvent and the solute as different dielectric regions were charges are distributed in a continuous way (Continuum Electrostatic methods) [227, 228]. Since most of the simulation time is spent screening the solvent configurations, the continuum solvent representations avoid its explicit sampling, which greatly decreases the computing time. However, in the Continuum Electrostatic (CE) methods, the detailed descriptions of solute-solvent interactions are lost [229].

In the works reported in this thesis, a combination of the Poisson–Boltzmann model and Metropolis Monte Carlo (MMC) simulations was used to determine the state of protonatable groups at a specific pH.

In the following sub-chapters, the CE methods will be generally explained to the reader.

2.3.1. Poisson-Boltzmann equation

The protein's electrostatic properties can be modeled using the Poisson–Boltzmann equation [230]. In this approach, the macromolecule is treated as an uniform low dielectric media with point partial charges at atomic positions and the solvent is represented by a high dielectric zone. The dielectric constant of the protein typically

ranges from 2 to 20 [231] and accounts for the electronic polarization and limited flexibility of the protein. A high dielectric constant (≈ 80) is usually assigned to the regions representing the solvent in order to account for the electrostatic configurational reorganization of the water molecules.

In a medium where the dielectric constant is not uniform and varies with the position, the Poisson equation can be used to treat the electrostatic interactions of the system. The Poisson equation (equation 2.14) relates the variation in the electrostatic potential $\phi(\boldsymbol{r})$ with the charge density $\rho(\boldsymbol{r})$ and the dielectric constant $\varepsilon(\boldsymbol{r})$:

$$\nabla \bullet \left[\varepsilon_0 \varepsilon(\boldsymbol{r}) \nabla \phi(\boldsymbol{r})\right] + \rho(\boldsymbol{r}) = 0 \tag{2.14}$$

where $\boldsymbol{r}$ is a given position vector, ε_0 the electric permittivity of vacuum, $\varepsilon(\boldsymbol{r})$, $\rho(\boldsymbol{r})$ and $\phi(\boldsymbol{r})$ are the dielectric constant, the charge density and the electrostatic potential at $\boldsymbol{r}$, respectively. However, when mobile counter-ions are present, the Poisson equation needs to be modified in order to account for the counter-ions redistribution in the solution as response to the electric potential, resulting in the linear Poisson-Boltzmann equation (equation 2.15):

$$\nabla \bullet \left[\varepsilon_0 \varepsilon(\boldsymbol{r}) \nabla \phi(\boldsymbol{r})\right] - \varepsilon(\boldsymbol{r}) \kappa^2(\boldsymbol{r}) \phi(\boldsymbol{r}) + \rho(\boldsymbol{r}) = 0 \tag{2.15}$$

where κ is the Debye-Huckel inverse length and it is defined as:

$$\kappa^2(\boldsymbol{r}) = \frac{2I}{\varepsilon k_{\mathrm{B}} T} \tag{2.16}$$

In equation 2.16, I is the ionic strength, k_B is the Boltzmann constant and T is the absolute temperature. The linear Poisson-Boltzmann is usually solved by numerical methods, such as the finite difference method [232].

2.3.2. Electrostatic binding free energy

A protonatable site in a protein can be any residue with a side chain able to capture and release a proton. The protonation of a site in a protein and the effect of inserting a titrable residue in a protein can be determined by a thermodynamic cycle like Figure 19.

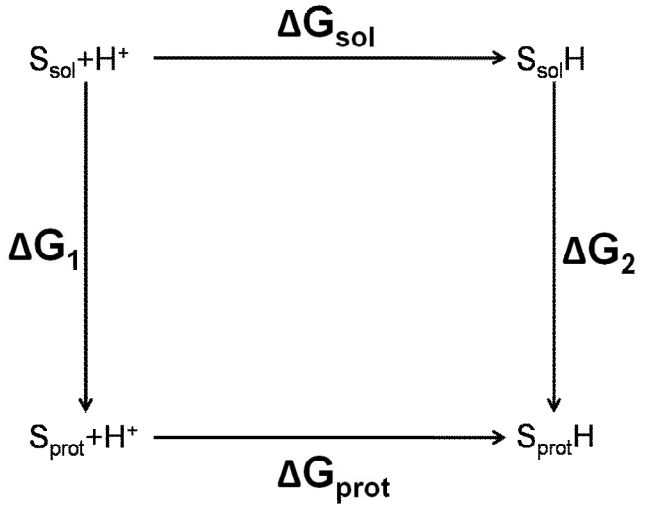


Figure 19- Thermodynamic cycle used to calculate the effect of inserting a titrable residue in a protein.

In this cycle, the S_{sol} and $S_{sol}H$ correspond to the deprotonated and protonated state of a specific site in solution, whereas S_{prot} and $S_{prot}H$ refer to the protonated and deprotonated states of the same site in the protein. ΔG_1 is the free energy difference obtained by moving the deprotonated site from the solution to the protein,

while ΔG_2 is the free energy difference resultant from moving the protonated site from the solution to the protein. ΔG_{sol} and ΔG_{prot} are the free energy difference of protonating a site in solution or in the protein, respectively. From the thermodynamic cycle, the ΔG_{prot} can be given by:

$$\Delta G_{\text{prot}} = \Delta G_{\text{sol}} + \Delta G_2 - \Delta G_1 \tag{2.17}$$

The binding constant of a proton inside a neutral protein is usually named intrinsic $\text{p}K_{\text{a}}$ ($\text{p}K^{\text{int}}$) and it is given by equation 2.18:

$$\text{p}K^{\text{int}} = \text{p}K^{\text{mod}} + \left\{ \frac{\Delta G_2 - \Delta G_1}{2.3 k_{\text{B}} T} \right\} \tag{2.18}$$

where $\text{p}K^{\text{mod}}$ is the $\text{p}K_{\text{a}}$ value for that site in solution and it can be inferred from experimental data.

The free energy terms related to the binding of protons to a site can be calculated using the Poisson-Boltzmann equation. By using the linear Poisson-Boltzmann equation, we can decompose the different energetic contributions in a sum of terms that depends not only in the individual sites but also on the pairs of sites [233]. If we considered a protein with N titrable sites, the binding state of this protein can be represented by the vector $\boldsymbol{n} = (n_1, n_2, n_3, \ldots, n_N)$ where n_i can only be 0 or 1, depending on whether site i is deprotonated or not. The standard free energy of protonation ($\Delta G(\boldsymbol{n})$), which is the free energy difference between a binding state $\boldsymbol{n}$ and a reference state (with all titrable sites deprotonated) can be defined as [234-239]:

$$\Delta G(\boldsymbol{n}) = -2.3k_{\mathrm{B}}T\sum n_i \mathrm{p}K_i^{\mathrm{int}} + \frac{1}{2}\sum_i \sum_{j\neq i} (n_i n_j + n_i z_j^0 + n_j z_i^0) W_{ij} \qquad (2.19)$$

where the value 2.3 stands for ln 10, the $\mathrm{p}K_i^{\mathrm{int}}$ is the intrinsic $\mathrm{p}K_{\mathrm{a}}$ of the site i, z_i^0 is the charge of the site i in the deprotonated form and W_{ij} is the direct electrostatic interaction between the sites i and j in the protein when $i \neq j$ [235, 238, 239]. $W_{i,j}$ can be written as the four combinations of the protonation state of sites *i* and *j* and the numbers 1 and 0 represents fully protonated and the fully deprotonated state of these sites, respectively:

$$W_{i,j} = w_{i,j}(1,1) - w_{i,j}(1,0) - w_{i,j}(0,1) + w_{i,j}(0,0) \qquad (2.20)$$

2.3.3. Binding equilibrium of protons

Upon the determination of $\Delta G(\boldsymbol{n})$, the probability of each binding state over a range of pH can be calculated, allowing the construction of the titration curve for a protonatable site as a function of pH. A binding state results from the combination of all protein's titrable sites, each one being whether protonated or deprotonated. The probability of occurrence of a particular state $\boldsymbol{n}$, $p(\boldsymbol{n})$, in thermodynamic equilibrium conditions, is given by:

$$p(\boldsymbol{n}) = \frac{\exp\left[-2.3z(\boldsymbol{n})\,\mathrm{pH} - \dfrac{\Delta G(\boldsymbol{n})}{k_{\mathrm{B}}T}\right]}{\sum_{\boldsymbol{n}'} \exp\left[-2.3z(\boldsymbol{n}')\mathrm{pH} - \dfrac{\Delta G(\boldsymbol{n}')}{k_{\mathrm{B}}T}\right]} \qquad (2.21)$$

where $z(\boldsymbol{n})$ is the total charge of state $\boldsymbol{n}$, $\Delta G(\boldsymbol{n})$ is the standard free energy of the binding reaction $0 \rightarrow \boldsymbol{n}$, k_{B} is the Boltzmann constant and T the temperature.

When using a method capable of calculating the binding free energy for each state $\boldsymbol{n}$ (like the method used by MEAD [234, 240]), the equation 2.21 can be directly used to sample the various protonation states of the titrable sites in the protein. However, and due to the exponencial dependence of the number of protonation states with the number of titrating sites, the direct application of equation 2.21 is not feasible for proteins with many sites [238], especially due to computer limitations. The alternative is to use a Metropolis Monte Carlo (MMC) method [241] to treat a significant number of sites with low sampling errors. This method uses a Metropolis criterion [218] to accept or reject the changes in the states (as previously described in section 2.2.8). The differences between states are changes in the protonation states of individual groups. In the end, and knowing the binding site populations for each site, the titration curve for a specific protonable site along the pH range can be constructed and the $\mathrm{p}K_{\mathrm{half}}$ (pH value at which the site is half protonated) of the site can be determined.

2.3.4. Proton tautomerism

In the CE methods, proteins are treated as rigid, which makes chemically equivalent protonable positions to become non-equivalent. A way to overcome this problem, is to allow the existence of alternative proton positions (proton isomerism), hereafter designated as tautomers [239]. In this case, each tautomeric site is decomposed into several non-tautomeric pseudo-sites (one per tautomer), which are then treated in the conventional way (for details associated to this formalism see [239]). When proton tautomerism is considered, the protonation state of the protein is represented by a vector $\boldsymbol{x}=(x_1, x_2, x_3, \ldots, x_N)$, where x_i is the protonation state of site i and x_i can have as many values as the total number of tautomers for that site (protonated and deprotonated). The standard free energy of protonation of state $\boldsymbol{x}$ can be written as (for details see [239]):

$$\Delta G(\boldsymbol{x}) = \sum_i g_i(x_i) + \frac{1}{2}\sum_i \sum_{j \neq i} g_{ij}(x_i, x_j) \tag{2.22}$$

Similarly to equation 2.19, in equation 2.22, the $\Delta G(\boldsymbol{x})$ is formed by an individual and a pairwise interaction term. g_i and g_{ij} are calculated after decomposing all titrable sites of the protein into non tautomeric pseudo-sites (for details see [239]).

Additionally, the probability of occurrence of a particular state $\boldsymbol{x}$, $p(\boldsymbol{x})$, in thermodynamic equilibrium conditions, is given by:

$$p(\boldsymbol{x}) = \frac{\exp\left[-2.3\, z(\boldsymbol{x})\,\mathrm{pH} - \dfrac{\Delta G(\boldsymbol{x})}{k_{\mathrm{B}}T}\right]}{\sum_{\boldsymbol{x}'} \exp\left[-2.3\, z(\boldsymbol{x}')\,\mathrm{pH} - \dfrac{\Delta G(\boldsymbol{x}')}{k_{\mathrm{B}}T}\right]} \tag{2.23}$$

where $z(\boldsymbol{x})$ is the number of protons bound to the sites in state $\boldsymbol{x}$.

In our CE calculations, tautomers were applied to all titrable residues, to the N- and the C-terminus, to water molecules, alcohols and to the nucleotide species. For the case of water and alcohols groups the existence of a charged form was forbidden.

2.3.3. Determination of pK^{int} and W_{ij}

The pK^{int} and the W_{ij} can be determined using the package MEAD [234, 240], which solves the Poisson-Boltzmann equation using the finite difference method [232, 242]. For the determination of the intrinsic pK_a (pK^{int}) and the interaction

term (W_{ij}), the dielectric constants used by the MEAD program were 80 and 20 for the solvent and the protein, respectively [239, 243]. However, since some of the ABC proteins studied are membrane proteins, the introduction of the membrane was essential. Originally, the MEAD package allowed the modeling of a lipid membrane represented as a single low dielectric slab of variable width. However, in our group a more detailed CE model for the lipid bilayer was developed, which defines the membrane as a low dielectric media of 41 nm formed by three different regions (see Figure 20). The first and third regions represent the hydrophilic lipid head group and they are 7.5 Å wide. The second region, with a thickness of 26 Å, characterizes the middle of the bilayer and mimetizes the hydrophobic tails of the lipids. Different values of dielectric constant were assigned for the different membrane slabs: for the first and third slabs of the bilayer the dielectric constant used was 20, while for the middle of the bilayer the dielectric constant used was 4.

Moreover, the complete transporters possess a water-filled transmembrane channel spanning from one side of the membrane to the other. The MEAD package also allows the introduction of a cylindrical shape "hole" through the membrane with a variable radius. However, and since the shape of the ABC transmembrane pathways does not resemble a simple cylinder, it was necessary to create a custom made hole in the membrane (see Figure 20). All points that are inside the channel and outside the protein interior are assigned with the solvent dielectric constant (ε_{sol} =80).

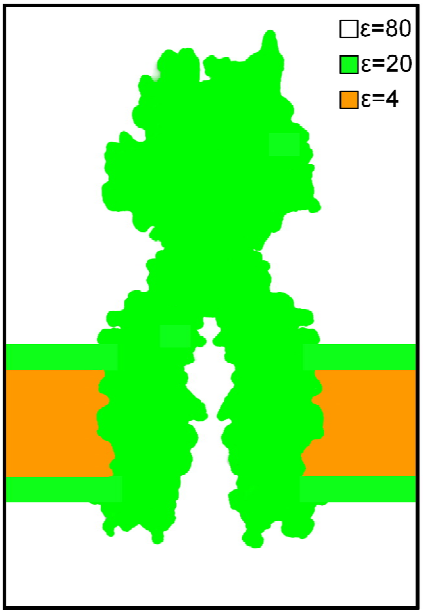


Figure 20- A schematic view of the model used to calculate the protonation states with the program MEAD for the Sav1866 exporter inserted in a lipid bilayer. The membrane is represented by three slabs of low dielectric media. The first and third slabs mimetize the hydrophilic lipid head groups of each leaflet of the bilayer (ε=20), and the middle slab (coloured in orange) represents the lipid hydrophobic tails (ε=4). In this model, the slabs representing the hydrophilic parts of the lipids are 7.5 Å wide, whereas the slab representing the lipid tails is 26 Å wide.

For all calculations, the temperature used was 300 K and a two step focusing procedure was applied for the Poisson–Boltzmann calculation. In the first step, the calculation was done in a grid of variable dimensions (depending on the protein size) with points spaced by 1 Å, followed by a focusing step on the site of interest, where the calculation was done in a smaller grid with 0.25 Å spacing.

3. Conformational changes in the MJ0796 NBD dimer

This work was published in the Journal of Physical Chemistry B: Ana S. Oliveira, António M. Baptista, and Cláudio M. Soares, (2010) "**Insights into the molecular mechanism of an ABC transporter: conformational changes in the NBD dimer of MJ0796**", J Phys Chem B, 114(16):5486-96.

3.1. Abstract

Despite the rapid advances in the study of ABC transporters, many fundamental questions linked to ATP binding/hydrolysis and its relation to the transport cycle remain unanswered. In particular, it is still neither clear nor consensual how the ATP energy is used by the nucleotide binding domains (NBDs) to produce mechanical work and drive the substrate translocation. The major conformational changes in the NBDs following ATP hydrolysis during the transport cycle and the role played by the conserved family motifs in harnessing the energy associated with nucleotide hydrolysis are yet unknown. Additionally, the way energy is transmitted from the catalytic to the membrane domains, in order to drive substrate translocation, is also a fundamental question that remains unanswered. Due to the high structure similarities of the NBDs architecture throughout the whole ABC family it is likely that the mechanism of ATP binding, hydrolysis and communication with the transmembrane domains is similar in all family members, independently of the nature of the transported substrate. In this work we focused our attention on the consequences of ATP hydrolysis in the NBDs, especially on the structural changes that occur during this process. For that we use molecular dynamics simulation techniques taking as a starting point the X-ray structure of the MJ0796 dimer from Methanococcus jannaschii. Several potential intermediate states of the ATP hydrolytic cycle are investigated, each consisting of different combinations of nucleotide bound forms. The results obtained allowed us to identify the conformational rearrangements induced by hydrolysis on the catalytic subunits, as well as the residues involved in this reorganization. The major changes are localized at specific regions of the protein, namely involving segments 11-19 and 93-124. Additionally, our results together with the knowledge of complete ABC transporters X-ray structures, suggest a possible NBD:TMD signal transmission interface.

3.2. Introduction

The ATP-binding cassette (ABC) transporters are a super family [244] of proteins that bind [74] and hydrolyze ATP [110, 153] to drive solute transport across membranes against the concentration gradient [86]. Transport substrates for these proteins include ions, sugars, amino acids, lipids, drugs, polypeptides and polysaccharides [86].

The members of this family play an important role in several physiological processes ranging from the import of nutrients to the export of toxic drugs from cells. Several members are associated with genetic diseases, such as cystic fibrosis [22, 245], and to multidrug resistance in bacteria, fungi, yeasts, parasites and mammals [15] (particularly the resistance of tumor cells to anticancer agents [246]). For this reason the study of this family has high biomedical interest and application.

Despite its diverse functions and the varied composition and structure of the transported substrate, several features are conserved across the ABC transporter family. For all members there is a basic functional unit formed by two cytoplasmic nucleotide binding domains (NBDs) and two transmembrane domains (TMDs) [86, 247, 248].

The TMDs form a transmembrane pathway for the substrate [86, 247, 248] and exhibit high sequence variability, presumably depending on the substrate transported. In contrast, the NBDs share high sequence and structure conservation between different family members. Several highly conserved sequence motifs involved in the ATP binding and hydrolysis can be identified in these domains [86, 247, 248]. This is the case of the Walker-A motif (also named P-loop), the Walker-B motif, the ABC signature motif (LSGGQ), the A-loop (an aromatic residue interacting with the adenine base of the bound nucleotide) [77], and two small sequences containing a conserved glutamine (Q-loop) [249] and a histidine (H-

loop). Functional studies of several ABC family members have provided evidence that the two NBDs interact and work as a functional unit [174].

As a consequence of the structure similarities found in the catalytic domains, it is plausible to assume a common mechanism of energy transduction powering the translocation, independently of the substrate transported. Although high-resolution X-ray structural data is available for the isolated NBDs (crystallized at several nucleotide-bound conditions [33-43, 47, 50-54, 91, 92]) and even for full-length transporters (vitamin B12 transporter BtuCD [57, 61], Sav1866 [58, 59], Hl1470/1 [63], $ModB_2C_2$ [60], MsbA [64], molybdate/tungstate transporter [250], methionine transporter [66], P-glycoprotein [65] and the maltose transporter $MalFGK_2$ [62, 142]), many fundamental questions about this family remain unanswered. In particular, the molecular details of the conformational changes due to ATP binding and hydrolysis, as well as their transmission to the TMDs, are yet unclear. To answer these questions, the elucidation of the atomic movements associated with the conformational changes that accompany substrate translocation is required.

MJ0796 is an ABC transporter found in the thermophilic Archaeon Methanococcus jannaschii [80, 188, 189] and belonging to the LolD transporter family [188]. The crystallographic structure for the NBD dimer [36, 51] has provided a clearer and currently accepted view of the interface of the NBDs. For the MJ0796 case, the mutation of the glutamate residue immediately following the Walker-B motif to a glutamine eliminated hydrolysis and allowed the crystallization of the dimer in the ATP bound state. This structure presents the NBDs dimerized in a nucleotide sandwich conformation having the two ATP molecules interacting with the P-loop residues of one monomer and the ABC signature motif of the other (see Figure 21). Both functional and crystallographic studies of several other ABC transporters indicate that this dimer interface represents the functional conformation in the intact transporters [51, 57, 251], and it is now consensually established that the MJ0796 structure reflects the competent conformation for transport [138].

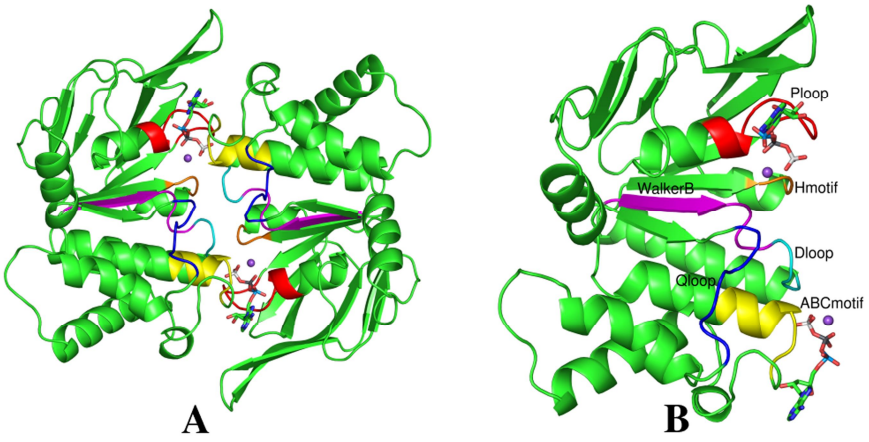


Figure 21- **A**) Structure of the MJ0796 mutant dimer with NaATP bound to catalytic site. ATP is represented with sticks and Na^+ as a purple sphere. Conserved sequence motifs are colored in red (P-loop), blue (Q-loop), yellow (ABC signature), magenta (Walker-B), cyan (D-loop), and orange (H-motif). **B**) The individual monomer with the conserved motifs colored as described in Figure 21A and labeled accordingly. These figures were generated with the program PyMOL [252].

By a still unknown molecular mechanism the binding/hydrolysis of ATP is supposed to induce rearrangements in the NBDs which would then be transmitted to the TMDs, allowing the transport of substrates. These types of molecular events are, in many cases, difficult to characterise by experimental studies with full atomic detail, due to technical difficulties. In this context molecular dynamics (MD) simulation methods [191] appear as a good option to address this type of questions. In the last decade, several studies using MD simulations have been reported, aimed at identifying the conformational changes in ABC family members [93-96, 162, 164, 165, 171, 173], some of them using the MJ0796 transporter as a model [87, 97, 172].

Campbell and Sansom [172] employed 10 ns MD simulations in order to identify the conformational differences between the MJ0796 dimer in three different nucleotide

bound states (ATP bound, ADP in one active site and ATP in the other, and the nucleotide free form). From this work they observed that the presence of ADP in one of the active sites promotes tighter (with higher number of H-bonds) protein-protein and nucleotide-protein interactions in the other site, demonstrating that there is close correlation between active sites in this dimer.

Jones and George [87] used molecular dynamics simulations and elastic network analysis applied to both the MJ0796 monomer and dimer in different bound forms (ATP-bound, ADP-bound and Free state for the monomer, and the ATP and ADP/ATP-bound form for the dimer). With the aim of mimicking the living conditions of this thermophile organism, all the MD simulations were performed at high temperature (358 K) and pressure (50 atm) conditions and the results for the dimer revealed a large asymmetrical movement located in the helical sub-domain of only one of the monomers, which the authors consider to be a consequence of the ATP hydrolysis. Additionally, while this manuscript was being finished, Jones and George [97], after extending the above described dimer simulations, observed that the rotation of the helical sub-domain relative to the catalytic one results in the opening of the ADP-bound active site. The authors claimed that this opening was probably sufficient to allow nucleotide exchange [97].

Moreover, in a recent work, Wen and Tajkhorshid [96], applied molecular dynamics techniques to study the detailed mechanism of dimer opening after ATP hydrolysis in the maltose transporter (MalK) NBD dimer. 50ns MD simulations of four possible ATP and ADP.IP (ADP plus Inorganic Phosphate) combinations were performed. After several tens of nanoseconds (30 ns in the fastest cases) it was possible to capture opening of one or both active sites (for all systems with the exception of the one with two ATPs) and to identify details associated to this process. The authors claim that a single ATP hydrolysis is enough to trigger the dimer separation.

A major difficulty with MD simulations of proteins is the limited sampling of the simulated ensemble. The limited time scales that can be currently simulated and the extrapolation of the general properties of a system by following the behaviour of only one molecule, as it is still often done nowadays, raise questions about their statistical significance. Questions like these do not have an easy and straightforward answer, since proteins have complex conformational energy landscapes with several local energy minima where the system can be trapped transiently or even permanently during the simulation [253]. Having this in mind, the use of MD replicates is becoming more frequent in protein MD simulation studies in order to minimize this problem. The idea is that, even if a simulation visits a few sub-states, each replicate will visit a different set of sub-states. So, the use of several replicates will help to overcome the sampling problem [254].

The prime goal of the work presented here is: to map the short time scale (<30 ns) conformational changes resulting directly from hydrolysis and to analyse their consequences in the context of the complete transporters. For that purpose we used MD simulation techniques, considering a considerable number of replicates, to study the MJ0796 NBD dimer in several intermediate states of the catalytic cycle, containing different nucleotides (ATP, ADP and the inorganic phosphate species) in the active site.

3.3. Materials and Methods

3.3.1. Starting Structure

The 1.9 Å resolution crystal structure of the mutant E171Q MJ0796 dimer (PDB code: 1L2T) [51] was used as the starting point for this work. In this structure the hydrolytic glutamate (residue 171) was mutated to a glutamine, producing a hydrolysis deficient protein that was stable in the dimer form [51]. So, the first task was to substitute this glutamine back to the corresponding acidic residue in order to recreate the wild type protein. The mutant was also crystallized with Na^{+}ATP instead of the Mg^{2+}ATP. For the simulations of the wild type the Na^{+} ion was replaced by Mg^{2+} in order to maintain a proper balance of the electrostatic charges in the active site. For the systems with ADP, the nucleotide coordinates were obtained by eliminating the γ-phosphate coordinates from the ATP molecule. For the systems that included the inorganic phosphate (IP), the positions of the IP atoms were similar to the positions occupied by the corresponding γ-phosphate group of ATP.

3.3.2. Protonation state of protonable residues

The protonation state of each individual group in the protein at a given value of pH (in this case pH=7.0) has to be specified prior to the MD simulations. We determined these protonation states using methodologies for studying the thermodynamics of proton binding described elsewhere [239, 255]. These methodologies use a combination of Poisson-Boltzmann calculations, performed with the package MEAD (version1.1.8) [234, 240, 256], and Metropolis Monte Carlo simulations, using the program PETIT (version1.3) [239]. According to these calculations (results not shown), performed on both the mutant X-ray structure [51] and the reconstructed wild-type, all lysines and arginines should be protonated, all

glutamic and aspartic acids should be deprotonated and the N-terminal should be in the charged state (NH_3^+). The C-terminus was considered neutral, since the crystal structure lacks some of the last residues for chains A and B. His141 is neutral (deprotonated at ND_1) while His204 is protonated. All free cysteines are protonated. The same set of protonation states was obtained for each monomer.

3.3.3. General setup of molecular dynamics simulation

All MD simulations were performed using the GROMACS 3.1.4 package [213, 257] and the 43A1 GROMOS96 force field [196, 199]. The SPC model [258] was used for water molecules. Crystallographic internal water molecules with a relative accessibility lower than 50% (determined using ASC [259, 260]) were included.

Starting from the X-ray structure, one system was initially prepared (see Figure 22A), corresponding to the wild type dimer with Mg^{2+}ATP in the two active sites. The protein model was solvated in a rhombic dodecahedral box, considering a minimum distance between the protein and box walls of 9 Å, containing 15016 water molecules. Two extra Na^+ ions were added to neutralize the **ATP/ATP** system, resulting in a system of 49752 atoms for the wild type. The ions were introduced using the program genion [213, 257], which replaced two solvent molecules by two monatomic ions at the positions with the most favorable electrostatic potentials.

The **ATP/ATP** system is the starting point for all simulations described here (except the ones with E171Q mutant). Five MD simulations were performed for this system, each 30 ns, in order to reduce the sampling problems (see section 3.2). All replicates were initiated with different sets of random velocities at the same temperature.

Every **ATP/ATP** system was energy minimized to remove excessive strain. First, we performed 5000 steps of steepest descent minimization with restraints (with a force constant of 1000 kJ mol^{-1} nm^{-2}) applied to all non-hydrogen atoms, followed by another 5000 steps of the same algorithm restraining only the C_α atoms. After the minimization step, we performed 50 ps of MD simulations with all the non-hydrogen atoms restrained (the force constant used was 1000 kJ mol^{-1} nm^{-2}) at the constant temperature of 300 K and constant pressure of 1 atmosphere, followed by 50 ps with position restraints only on C_α atoms. Unrestrained production simulations started from these relaxed conformations.

Table 3- Summary of all the simulations performed.

[a]- All simulations have in the active site, both the nucleotide species and the magnesium ion as a cofactor, with the exception of the mutant simulations, which have a sodium ion.

Simulation[a]	Time (ns)	Nucleotide in the active site		Temperature (K)	Pressure (atm)	Number of replicates
1) **ATP/ATP**	30	ATP	ATP	300	1	5
2) **ATP/ADP**	30	ATP	ADP	300	1	5
3) **ATP/ADP.IP**	30	ATP	ADP+IP	300	1	5
4) **ADP.IP/ADP.IP**	30	ADP+IP	ADP+IP	300	1	5
5) **ADP/ADP**	30	ADP	ADP	300	1	5
6) **ATP/ATP_358K**	30	ATP	ATP	358	50	5
7) **ATP/ADP_358K**	30	ATP	ADP	358	50	5
8) **ATP/ADP.IP_358K**	30	ATP	ADP+IP	358	50	5
9) **E171Q-ATP/ATP**	5	ATP	ATP	300	1	10

After 2 ns of unrestrained MD simulations of **ATP/ATP** at 300 K and 1 atm, several new systems were prepared by replacing the original nucleotides with ADP or ADP+IP (see Table 3 and Figure 22A). Four systems were thus studied: **ATP/ADP**, **ATP/ADP.IP**, **ADP/ADP** and **ADP.IP/ADP.IP**. For each system, water molecules were replaced by ions in order to achieve a total zero charge.

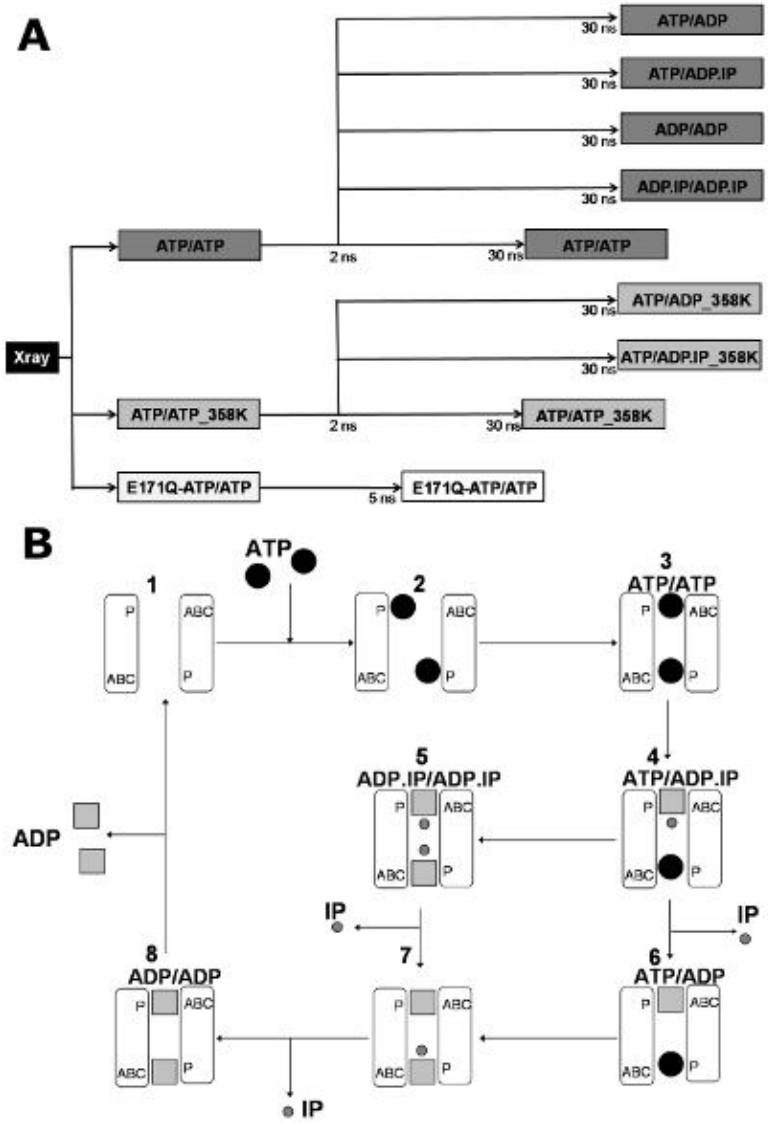


Figure 22- **A-** Outline of the MD simulations performed. Starting from the X-ray structure, the wild type dimer with ATP and cofactor in the active site was simulated both for the 300K/1atm (**ATP/ATP**) and for the 358K/50atm (**ATP/ATP_358K**) conditions. Then, after 2 ns of simulation, several other systems with different nucleotide bound states were prepared. Five replicates, of 30 ns each, were done for all the wild type systems reported. This figure represents the procedure used for only one replicate. Parallel to the wild type dimer, the mutant E171Q with ATP in the active site (**E171Q-ATP/ATP**) was also simulated. For the mutant systems, 10 replicates 5 ns long were performed. **B-** The Processive Clamp Model [136], one of the proposed mechanisms for the ATP catalytic cycle in the NBDs. The ATP is represented by black spheres, ADP by grey squares and IP by small grey spheres. P and ABC represent the P-loop residues and the ABC signature motifs respectively. The active pocket is formed by the P-loop residues of one NBD and the ABC signature of the other.

The time step for integrating the equations of motion was 0.002 ps. A cutoff of 14 Å was used for van der Waals interactions and a smooth particle mesh Ewald method [205] was used for long range electrostatic interactions beyond a 9 Å cutoff. The neighbor lists were updated every 10 steps. The SETTLE algorithm [217] was used for keeping water molecules rigid and the LINCS algorithm [216] was employed to keep all remaining bonds at their equilibrium lengths. The Berendsen coupling algorithm [261] was used to keep constant temperature and pressure conditions through the simulations. The pressure coupling constant used was 0.5 ps and compressibility was set to 4.5×10^{-5} bar^{-1}, while the heat coupling constant was set to 0.1 ps. There was a separate coupling of solutes and solvent baths [261].

After nucleotide exchange, both the **ATP/ADP** and **ADP/ADP** systems were submitted to 100 ps of MD simulation with all non-hydrogen atoms restrained (the values for the force constant, temperature and pressure were the same as for the **ATP/ATP** systems), followed by 100 ps of Cα position restrained MD.

In order to accommodate the IP species, the **ATP/ADP.IP** and **ADP.IP/ADP.IP** systems were submitted, first to a minimization step (with a procedure similar to the **ATP/ATP** systems), then to 5 ps of fully unconstrained (no LINCS) MD simulation

(using a time step of 0.0005 ps), and finally to restrained MD simulations (100 ps with restraints on all non-hydrogen atoms followed by 100 ps with restraints on C_α atoms). This procedure allowed for a slight rearrangement of the nucleotide without significant structural consequences on the protein.

In order to compare with the results obtained by Jones and George [87] at high temperature and pressure, some simulations were performed at 358 K and 50 atm. The starting structure for the ATP-bound high temperature/pressure systems (**ATP/ATP_358K**) was the same as for the **ATP/ATP** wild-type systems and the initialization protocol was similar to the one described above. Again, and similarly to the 300K/1atm **ATP/ATP** simulations, after 2 ns of unrestrained simulations several new systems were prepared by replacing the original nucleotides with ADP or ADP+IP (all the details for these simulations are listed in Table 3 and Figure 22A).

Some control simulations were also performed for the E171Q mutant with ATP-bound in the active sites (**E171Q-ATP/ATP**), in order to clarify the structural effect of the mutation. All replicates for the control system started from the X-ray structure with different sets of random velocities. For the **E171Q-ATP/ATP** systems, 10 replicates each 5 ns long were performed. The details for these simulations are again listed in Table 3 and Figure 22A.

For details related to the ATP, ADP and IP parameterization see the Supporting Information section 8.1.

3.3.4. Data analysis

All average structures were computed by neglecting the first 20 ns of simulation. Least-squares fitting was applied to the Cα atoms using the X-ray structure as reference.

The secondary structure assignment was performed using the DSSP program by Kabsch and Sander [262]. To determine the percentage of secondary structure loss relative to the X-ray structure, the regular secondary structure classes considered were the α-helix, the β-sheet, the 3_{10} helix and the β-bridge (DSSP classification). Only the residues that remain in the same secondary structure class present in the X-ray structure are counted.

The atomic positional deviations were calculated by comparing the different nucleotide bound states to the **ATP/ATP** systems within the same replicate, and averaging these differences over all replicates. This procedure is based in the "subtraction technique" introduced by Ciccotti et al [263], but in this case applied to a much larger timescale. In accordance to the "subtraction technique" [263], an initial unperturbed state is used to start two new simulations: one with the perturbation and the other unperturbed (simply continuing the initial trajectory). In this case, the response of the system to the introduction of a perturbation can be measured from the direct comparison of a perturbed and an unperturbed trajectory. Applying the "subtraction technique" rationale to this work, we will have the **ATP/ATP** simulations as the unperturbed state and the replacement of ATP by other nucleotide as the perturbed state (**ATP/ADP**, **ATP/ADP.IP**, **ADP/ADP** or **ADP.IP/ADP.IP**). In this way, the response of a property to the perturbation can be directly computed as the difference between the two states (e.g. the replicate i of the **ATP/ADP** system is compared to the replicate *i* of the **ATP/ATP** system). This comparison within replicates allows the elimination of the structural differences arising from natural variation between replicates.

3.3.5. Catalytic cycles for the ABC NBD dimer

At this moment, there are three proposed models [135] reported in the literature for the ABC NBD ATP hydrolysis: the Processive Clamp, the Alternating Sites and the Constant-Contact models. The first two models differ on the number of ATP

molecules required to open the dimer, while the third model suggests that there is no dimer separation and the rotation of the helical sub-domain is enough to allow nucleotide exchange. In this work, each of the above described nucleotide combinations correspond to different states in a possible ATP catalytic cycle, which are the situations we are simulating, irrespectively of the model considered. Some of the systems studied in this work are unique to the Processive Clamp Model [136] shown in Figure 22B (states **ADP.IP/ADP.IP** and **ADP/ADP**), while others are also common to both the Processive Clamp and the Alternating Sites models (states **ATP/ATP** and **ATP/ADP.IP**). The **ATP/ADP** system is common to all three proposed models.

3.4. Results

3.4.1. Structural differences between the E171Q mutant and the wild type dimer

Since the X-ray structure chosen as the starting point for this study is a mutant (**E171Q-ATP/ATP**) and thus may not represent the exact details of the natural physiological dimer, we started this work confirming the structural equivalence between the mutant and the wild type (**ATP/ATP**). Ten molecular dynamics simulations of 5 ns each were performed for both systems in order to characterize the structural differences between them; this was done by calculating the root mean square fluctuations (RMSF) and the root mean square deviations (RMSD) to the X-ray structure for each residue, as well as the average three-dimensional structure (see the Supporting Information section 8.2). From the results we could conclude that both the wild-type and the mutant have a similar behaviour during the MD simulations and that the residues displaying higher RMSD and RMSF are the same in the two systems. This led us to consider the two systems structurally similar and validate the use of the mutant X-ray structure for studying the wild type dimer.

3.4.2. Structural differences during the catalytic cycle

A simple measure of the overall stability and structural drift of the protein can be obtained by plotting the RMSD of the Cα atoms as a function of time (Figure 23).

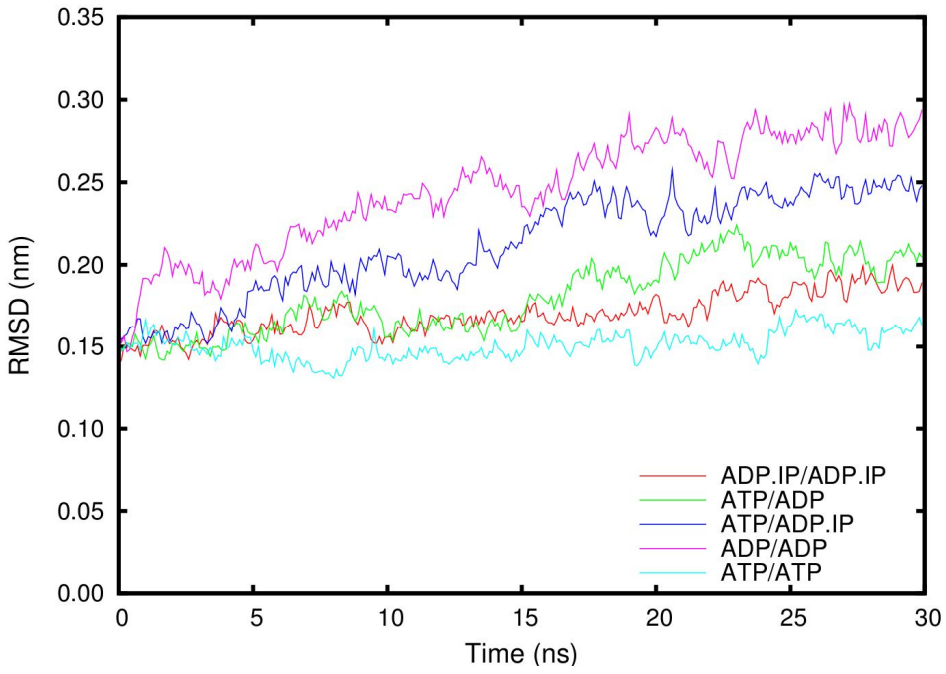


Figure 23- Time evolution of the Cα RMSD relative to the **ATP/ATP** average (least squares positional fitting made to this same average structure). The global **ATP/A**TP average structure was obtained over the last 10 ns for all of the 5 replicates after superimposition to the X-ray structure. RMSD values were averaged from all 5 replicates for each state. Additionally, each point in the plot represents the average RMSD value for 100 ps.

As can be observed, the RMSD increases until it reaches the global value of about 0.16 for the **ATP/ATP**, 0.18 nm for the **ADP.IP/ADP.IP,** 0.21 nm for the **ATP/ADP,** 0.24 nm for the **ATP/ADP.IP** and 0.28 nm for the **ADP/ADP.** In particular, no significant drifts in this measure were observed in the last 10 ns. The RMSD evolution (relative to the X-ray structure) for all 5 replicates simulated for all the states reported in this study can be observed in the Supporting Information section 8.3.

For all systems, the tertiary structure of the dimers was monitored during the simulation time and a residual secondary structure loss (9%) from the X-ray structure was observed within the simulated time scale.

The dimer stability was examined for all systems by monitoring several properties along the simulation, namely the number of NBD_1-NBD_2 H-bonds, the radius of gyration and the solvent accessible surface (results not shown). All these properties were kept approximately constant during the whole 30 ns of simulation for all the systems.

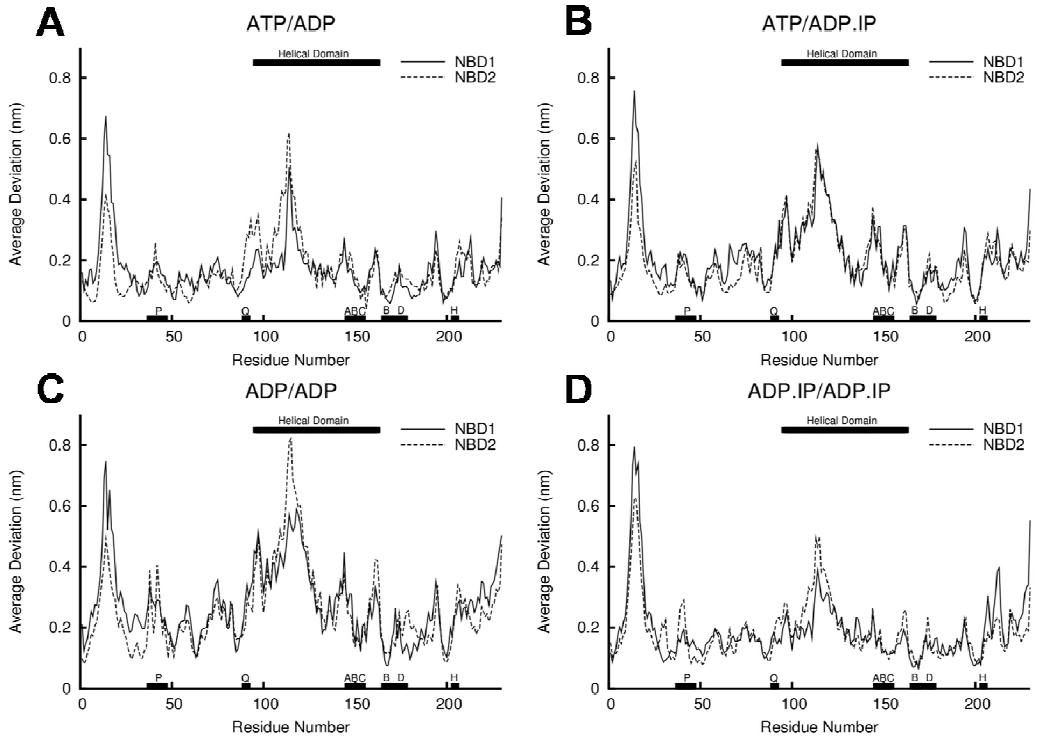


Figure 24- Average Cα Positional Deviation (over the last 10 ns of the simulations) relative to the average **ATP/ATP** conformation within replicates for: **A) ATP/ADP**, **B) ATP/ADP.IP**, **C) ADP/ADP** and **D) ADP.IP/ADP.IP**. The deviation values displayed for the different nucleotide bound states were computed by comparing the average structure for each replicate with respect to the average **ATP/ATP** conformation for the same replicate. Then these individual differences were averaged over all

replicates. The graphs show the position (in the polypeptide chain) of the ABC structural family motifs and some relevant sub-domain regions: P-loop (P), Q-loop (Q), ABC signature motif (ABC), Walker-B motif (B), D-loop (D), H-motif (H) and Helical Domain (residue 96-162).

In order to identify the residues involved in the conformational changes occurring during ATP hydrolysis, the Cα positional deviation relative to the **ATP/ATP** system was determined as a function of the residue number for the last 10 ns of simulation for all systems. As seen in Figure 24, in all the cases we can observe higher average Cα deviation values for residues 11-19 and 93-124, relative to the **ATP/ATP** simulations, while the remaining residues show more or less similar behaviour upon hydrolysis. Moreover, the major conformational changes observed are not restricted to the areas surrounding either the active site or the hydrolysed nucleotide. An example of this observation is the conformational change for the residues 93-124 located in the helical sub-domain, quite far apart from the active site, where the actual ATP hydrolysis occurs. This fact may suggest that the mechanism of energy transduction from hydrolysis is a very complex process and it is not limited to the areas surrounding the nucleotide binding pockets. Other studies, both experimental [264] and theoretical [87], have also suggested that the region equivalent to residues 93-124 in MJ0796 may be coupled to the nucleotide changes in the active sites.

Another interesting observation is the fact that some of the conserved sequence motifs that bind to the nucleotides (Walker-B, ABC signature motif and the D-loop) do not show relevant differences in terms of positional deviation values between the ATP- bound state and all the other states. Only the P-, H- and the Q-loop residues show slightly higher positional deviation values when in the presence of ADP (the NBD_2 of Figure 24A and Figure 24C) or ADP+IP (Figure 24B). This suggests that, upon binding of the nucleotide (whether ATP or ADP), the residues of the ABC family motif are stabilized, adopting an approximately fixed conformation. From the comparison of the available X-ray data (ATP-bound dimer [51] and ADP-bound monomer [36]), it was observed that the ABC signature motif is the region

undergoing the most relevant movement upon hydrolysis. This does not happen in our simulations, and in fact, the ABC motif is one of the regions presenting lower positional deviation values relative to the ATP-bound state.

As evidenced from Figure 24B, and despite the hydrolysis being only located in one of the binding pockets, the monomer behaviour is more or less symmetric, which again reinforces the idea that the communication between the two monomers is a very complex mechanism. For the ADP-bound systems (Figure 24A and Figure 24C), the conformational change previously observed for residues 93-124 (helical sub-domain) is located in the same region as the one presenting largest differences between the different nucleotide bound states characterised experimentally [36, 51]. Based on the comparison between the ATP-bound crystallographic dimer [51] and the ADP-bound crystallographic monomer [36], Smith and co-workers reported a rigid body rotation of the helical sub-domain. Moreover, similar conformational changes were reported for the ADP-bound structure of MJ1267 [35]. Although there is a good correlation between the location of the conformation change observed in our simulations and the experimental data, the details for the helical sub-domain movement are significantly different, especially in the ABC signature region, as noted above. Nevertheless, it is important to stress that the X-ray conformational differences between states were obtained by comparing the ADP-bound monomer structure to the ATP-bound dimer structure.

Each curve in Figure 24 is the average of five same-replicate pairs of simulations, as indicated in the legend. As explained in the Materials and Methods section, this intends to take into account eventual correlations between simulations corresponding to the same replicate. Nonetheless, the average of all 25 possible pairs of simulations produces similar results, as shown in Figure 48 of the Supporting Information section 8.4. This suggests that no significant correlations remain after such long simulation times. Furthermore, the fact that the regions

identified as exhibiting hydrolysis-induced changes do not depend on the averaging method is a good indication of their reliability.

Mapping the positional differences in the average three-dimensional structure for every system, we can identify the spatial location of the residues that undergo larger displacement during a possible catalytic cycle (Figure 25).

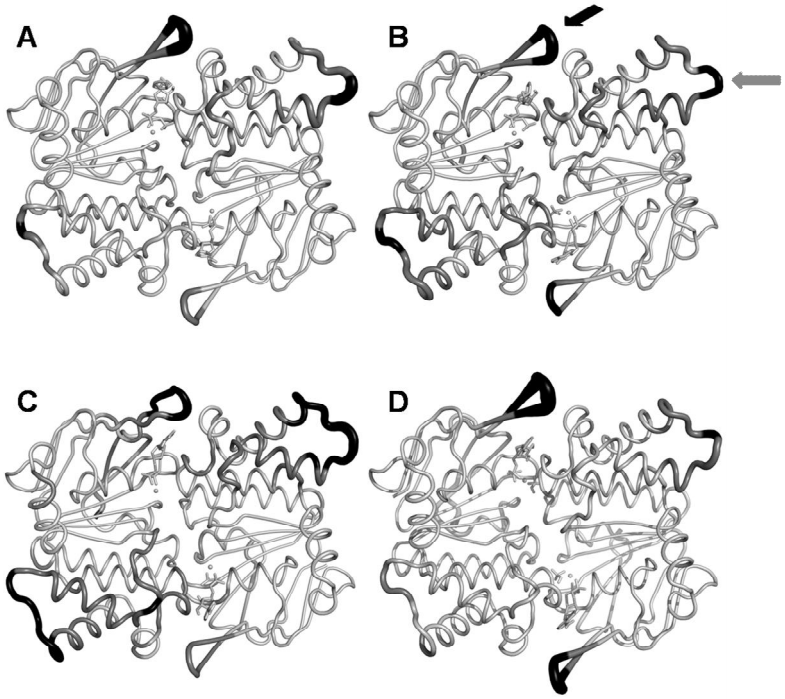


Figure 25- Average three-dimensional structure of the wild type systems with different nucleotide bound species. The average structures were determined from the last 10 ns of simulations over all the replicates. The regions presenting high deviations relative to the average **ATP/ATP** structure are represented with a higher main chain radius, while the ones with lower deviation are thinner. The regions with the positional deviations above 0.50 nm are coloured in black while the residues displaying differences in position between 0.30 and 0.50 are coloured in grey. The protocol for obtaining the positional deviation values is the same as in the legend of Figure 24. All averages were fitted to the X-ray structure. The protein is rendered as a main chain cartoon; the ATP, ADP and IP are represented with sticks while the sphere corresponds to the Mg^{2+} ion. All figures were generated with

the program PyMOL [252]. The systems represented are: **A**) **ATP/ADP**, **B**) **ATP/ADP.IP**, **C**) **ADP/ADP**, **D**) **ADP.IP/ADP.IP**.

Two regions with high deviations are observed. The region showing the largest structural rearrangements induced by the hydrolysis and IP release corresponds to residues 93-124, located in an external and flexible region of the dimer, called the helical sub-domain (grey arrow in Figure 25). Based on previous experimental [35, 36, 86] and theoretical works, some authors have already suggested that this region rotates outward after γ-phosphate cleavage, presumably to facilitate ADP release from the active site. Nevertheless, it is important to remember that most of the experimental studies compare ADP-bound monomers and ATP-bound dimer mutants [35, 36]. The second set of altered residues range from 11-19, and are located in the exterior loop connecting two β-sheets S_1 and S_2 (black arrow in Figure 25). This loop is positioned just above the adenine ring of the nucleotide. An essential conserved tyrosine (Y_{11},named A-loop [77]) is located in this region. In our simulation, and similarly to what was suggested by other authors [77] this residue interacts with the adenine base of nucleotide during the hydrolytic cycle by π–π stacking (data not shown) [265].

At this point it is important to refer a theoretical study published by Jones and George where they perform 22 ns MD simulations for the MJ0796 dimer in different ligand bound states (ATP-bound and ATP/ADP-bound) at high temperature (358 K) and pressure (50 atm) [87]. The use of high temperature and pressure conditions was justified by the authors on the argument that *Methanococcus jannaschii*, the organism where this protein is found, lives in these conditions. In this work the authors identified a very large movement (rotation of about 13 Å of the $\alpha_{3\text{-}4}$ loop) of one of the helical sub-domains in the ATP/ADP-bound system, which they consider to be a consequence of the hydrolytic process. This same movement was not observed in the simulations of the ATP-bound state. After comparing our results

with the ones presented by Jones and George a curious observation was made: although the set of residues representing the largest conformational change upon hydrolysis is more or less in the same region (in the helical sub-domain) for both works, the movement reported by Jones has much larger amplitude than the one we observe. Despite the differences in amplitude, our results are qualitatively similar to those observed by Jones and George [87] and are clearly distinct from the ones inferred from the X-ray analysis [36, 51]. However, while Jones and George [87] observed that ATP hydrolysis was associated with both the rotation of this helical sub-domain and with the loss of a Q_{90}-Mg interaction, we found no such correlations in our simulations. In fact, the direct coordinating Q_{90}-Mg interaction (0.2-0.25 nm) is never present in our simulations and the distance between the Q90 and the magnesium is almost always above 0.4 nm, even in the **ATP/ATP** simulations. In the Malk dimer, Wen and Tajkhorshid [96] observed the formation of a direct interaction between the Mg and the Q-loop glutamine in one binding site, which remained intact in the presence of ATP but, similarly to the Jones and George [87] findings, was lost upon replacement by ADP-IP.

To see the effect of the high temperature and pressure on this system, we carried out five 30 ns MD simulations of the ATP-bound (**ATP/ATP_358K**), ATP/ADP.IP-bound (**ATP/ADP.IP_358K**) and the ATP/ADP-bound (**ATP/ADP_358K**) systems using the same conditions of temperature (358 K) and pressure (50 atm) reported by Jones and George [87] but using the GROMOS force-field. The protein stability was carefully examined during these simulations by monitoring several system properties, namely, the number of H-bonds formed between the two monomers, the radius of gyration, the solvent accessible surface, the native α-helices H-bonds and the secondary structure (data not shown). For these high temperature and pressure simulations we observed native secondary structure loss for all systems (**ATP/ATP_358K**, **ATP/ADP_358K** and **ATP/ADP.IP_358K**) without exception. The loss of secondary structure (relative to the X-ray) is 18% in the 358K/50atm simulations, but only 9% in the 300K/1atm ones.

The average RMSD and standard deviation values from the X-ray structure (after X-ray fitting) over the last 20 ns of simulation observed for NBD_1 and NBD_2 in the **ATP/ATP_358K** system was respectively 0.251±0.02 and 0.306±0.01 nm, while the one previously reported for the ATP-bound isolated monomer [87] was 0.221±0.03 nm.

We cannot rule out the possibility that this structural deviation from the crystal structure indicates unstable behaviour as a consequence of the fact that the high temperature and pressure conditions used here are way above the conditions for which the GROMOS force-field is developed and parameterised. Nevertheless, we feel that this data is worth reporting here. Moreover, in one of the five replicates of the **ATP/ATP_358K** system (where no hydrolysis occurred) we could also observe the existence of the same large scale movement in the helical sub-domains considered by Jones and George to be a consequence of nucleotide hydrolysis. A similar helical sub-domain movement was also observed at the end of the trajectories in two replicates for the **ATP/ADP_358K** and in three replicates for the **ATP/ADPIP_358K.**

At this stage, we cannot clearly state if this large scale movement observed here in the high temperature/pressure simulations is a reality or an artefact, given the differences between our results and the ones obtained by Jones & George [87]. In any case, although the high temperature/pressure conditions typical for this thermophile may facilitate the occurrence of some molecular events, the protein is fully functional at room conditions [80].

3.4.3. A possible exit way for Inorganic phosphate

It is proposed in the literature [51] that the presence of a charged molecule such as IP near an also highly charged nucleotide (in this case ADP) can be at the origin of

the spontaneous dissociation of the dimer with posterior ADP release. Based on the results obtained in this work by the analysis of IP-containing systems, the introduction of an additional negative charge in the nucleotide binding pocket has very little effect on the overall three-dimensional structure of the dimer; this is evident from the observation of the average structure for **ADP.IP/ADP.IP** (Figure 25D) and **ATP/ADP.IP** (Figure 25B) and the low RMSD value obtained for the IP-containing systems (Figure 23). Moreover, a somewhat curious observation was the low RMSD value obtained for the **ADP.IP/ADP.IP** system (Figure 23). A possible explanation to this observation is to consider that, although the charged nucleotide ADP is present, the IP molecule arranges itself in order to reduce repulsion and optimize the non-bonded interactions, taking the place of the γ-phosphate (in the tail of ADP), mimetizing the ATP molecule.

Despite the fact that no drastic changes in the tertiary structure for the IP containing systems were observed, the contact area between the monomers is slightly reduced (about 12% of contact area lost when compared with the **ATP/ATP** systems) when both active sites contain simultaneously ADP and IP. At this point, an interesting question can be placed: after ATP hydrolysis, how does IP exit out of the active site?

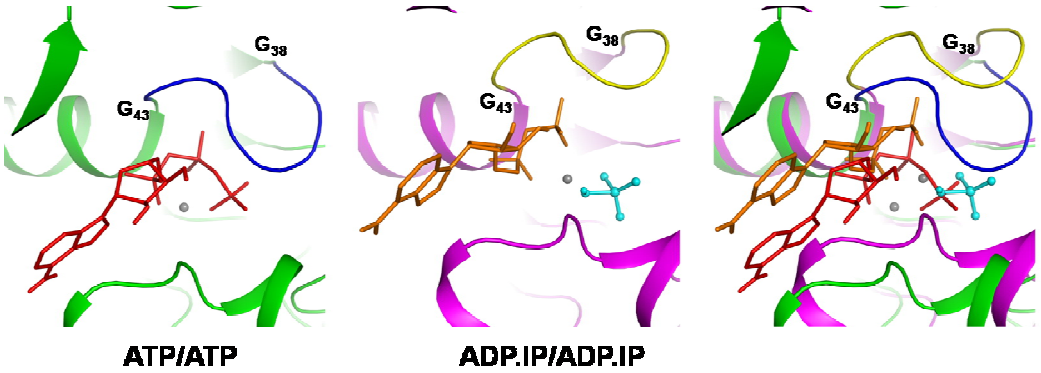


Figure 26- Detail view of the movement of loop 38-43 in replicate 2, which leads to the solvent exposure of the IP species. The left hand side shows the average structure obtained for replicate 2

(over the last 10 ns of simulation) for the ATP-bound system (**ATP/ATP**). In this case, the backbone is coloured in green and the loop 38-43 in dark blue. ATP is shown as red sticks whereas the magnesium ion is represented as a gray sphere. The central picture represents the average structure, also obtained for replicate 2 (over the last 10 ns of simulation), for the ADP.IP-bound system (**ADP.IP/ADP.IP**). Now, the protein main chain is coloured in magenta while the loop 38-43 is yellow. ADP is depicted by orange sticks whereas the IP species is represented by cyan ball and sticks. The right figure corresponds to the superimposition (after C-α fitting) of the two left and central pictures. The color code is identical to the two previous figures.

A very interesting observation in the IP-containing systems is the movement of the loop 38-43 which leads to the opening of a cavity and thereby exposes the IP to the solvent (see Figure 26). The set of residues 38-43 is part of the highly conserved P-loop sequence motif. For the **ADP.IP/ADP.IP** systems this movement was observed for all the replicates at least in one of the active sites. Additionally, for the **ATP/ADP.IP** systems this occurrence was also observed in two of the five replicates. In Figure 26 (right figure) we superimpose the ATP-bound and the **ADP.IP/ADP.IP**-bound average for replicate 2 as an example of the above described movement. As can be seen, the loop in question moves away from the active site leaving the IP totally exposed to the solvent. This may allow it to freely diffuse from the active site, something that did not occur in the time-scale of our simulations, but may occur in longer time-scales.

Additionally, in order to determine whether the IP species is more accessible to the solvent than the γ-phosphate group from ATP, the solvent accessible surface (SAS) was determined for all systems containing whether ATP or ADP plus IP (Table 4):

Table 4- Average SAS of either the IP species or the γ-phosphate group of ATP, for systems containing ATP or ADP+IP. [a]- The averages are computed using the last 10 ns of all replicates. To determine the SAS for the ATP γ-phosphate group, we consider the phosphate atoms plus all the oxygens bound to it (both the O_3PB and the terminal oxygens -see Figure 44C in the Supporting

Information section 8.1-). [b]- Errors (in parenthesis) are shown as +/-2 standard deviations of the average, corrected for time correlations using standard techniques [266].

Simulation	SAS [a] (+/-2*stdev) [b]	
	NBD1 (nm^2)	NBD2 (nm^2)
ATP/ATP	0.041 (0.002)	0.045 (0.009)
ATP/ADP	0.076 (0.015)	-
ATP/ADP.IP	0.142 (0.027)	0.158 (0.016)
ADP.IP/ADP.IP	0.152 (0.006)	0.199 (0.036)

As can be observed, whenever the IP species are present, the SAS always presents an increase when compared to the γ-phosphate group of the **ATP/ATP** system. In the case of the **ATP/ADP.IP**, the presence of the inorganic phosphate in one binding pocket induces conformational changes in the dimer interface which are eventually transmitted to the other ATP-bound site and by so exposing the γ-phosphate group of ATP to the solvent.

Besides IP release, the question on how ADP leaves the active site remains open. Wen and Tajkhorshid [96] reported clear dimer opening upon ATP hydrolysis in their simulations of the MalK dimer in time scales ranging from 30 to 50ns. This opening can be the first step for ADP release. In the simulations reported here, dimer opening has not occurred in any of the replicates within the simulated time scale, although the contact area between monomers decreases by about 12% (see above). Despite the large differences between MalK and the MJ0796 protein (MalK has an extra regulatory domain), and the different simulation conditions (different force fields and different water molecules), one cannot rule out that our system will eventually open in a longer time scale.

3.4.4. Localization of the major conformational changes in the full-length transporter: inter-domain interfaces

It is generally accepted that the conformational changes induced during the ATP hydrolytic cycle are transmitted to the membrane domains by the direct contact of several residues located at the NBD-TMD interface. Previous studies performed in different ABC members, both experimental [86 , 267] and theoretical [87, 173], have suggested that the inter-domain interface may consist of the helical sub-domain, which is essentially the same set of residues identified here as those presenting the largest conformational rearrangements during nucleotide hydrolysis and IP release. This sub-domain is one of the most structurally diverse among this family, which can be due to the high sequence variability of the TM domains it contacts with [173]. In order to have an idea of the position of residues 11-19 and of the helical sub-domain in the overall complex, we performed the structural alignment of the average structure for the **ATP/ATP** system with the X-ray structure of the B_{12} full length importer from *E. Coli* [57].This allowed us to confirm the direct contact of the helical sub-domain residues with the membrane spanning domains of the complete transporter (see Figure 27). The results obtained for our MD simulations (where the helical sub-domain presents the most pronounced conformational changes between different bound nucleotides systems) suggest a fundamental role for this segment in the transmission of the structural changes from the active site to the transmembrane domains.

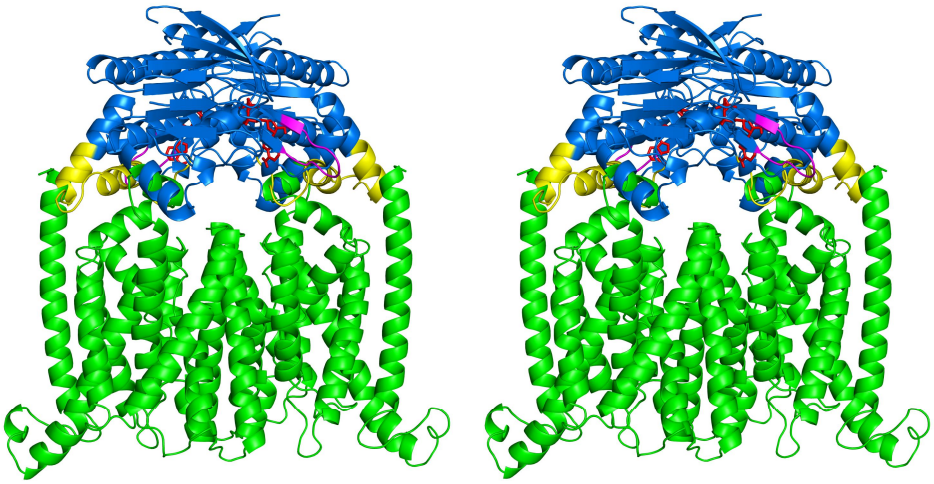


Figure 27- Stereo view of a hypothetical model for the complete MJ0796 transporter constructed after structure alignment of the MJ0796 NBD (ATP bound average over the last 10 ns of simulation) dimer with the NBDs from the vitamin B_{12} complete X-ray structure (NBDs+TMDs). The vitamin B_{12} transmembrane domains are colored in green, while the MJ0796 NBDs are colored in dark blue with the regions presenting the biggest backbone movements, residues 11-19 and 93-124, highlighted in magenta and yellow, respectively. The ATP species are represented as red sticks and the magnesium ion as a grey sphere. This chimera was only built for visualization purposes and in order to identify the spatial position of the some specific set of residues in the overall complex.

Other experimental works have also suggested a possible critical role for the X-loop (which is a small “TEVGERG” sequence motif conserved in ABC exports only) in inter-domain communication [268, 269], but we have not observed any relevant change in our simulations upon hydrolysis.

3.5. Concluding Remarks

Nowadays, it is generally accepted that ATP binding and hydrolysis is needed in ABC transporters to drive the translocation of substrates across the membrane. During one catalytic cycle ATP hydrolysis would induce conformational changes in the catalytic subunits and modulate NBD-TMD interactions, coupling the conformational changes to the transport of substrates.

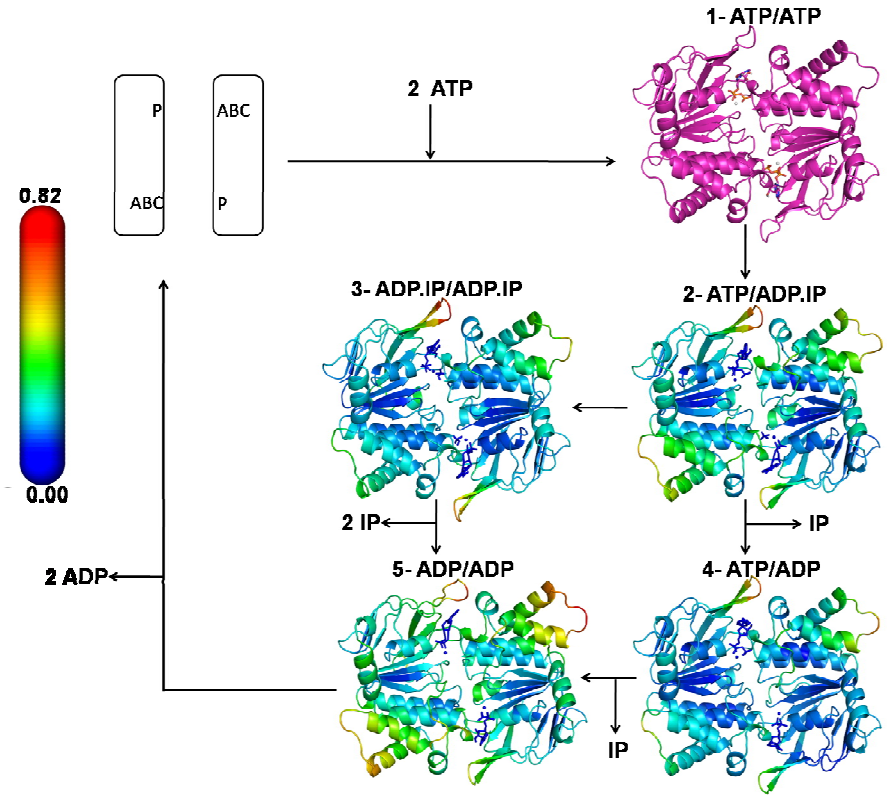


Figure 28- A possible mechanism for the ATP hydrolysis cycle for the NBDs of ABC transporters with the conformational rearrangements mapped in the average three-dimensional structures of each step. The average structures were determined from the last 10 ns of simulations over the replicates (using the X-ray structure as reference). The resting state of the dimer (without bound nucleotides in the active sites) is schematically represented by white rectangles, with the “P” and “ABC” labels

representing the P-loop residues and the ABC signature motifs, respectively. The average structures are coloured according to the positional deviation gradient. The Cα positional deviation values were calculated between the average conformations for each replicate (obtained after X-ray fitting) with different nucleotide bound forms and the same replicate averages structure for the **ATP/ATP** (the magenta figure) system. The protein is represented as a cartoon while the ATP, ADP and IP are represented as sticks and the sphere corresponds to the Mg^{2+} ion. It should be noted that there is no direct correspondence between the states in Figure 22B and in Figure 28. Figure 28 shows the systems explicitly studied in this work, which are only some of the states in the Processive Clamp Model of Figure 22B.

In this work, several intermediate states in the ATP hydrolytic cycle of the MJ0796 NBD dimer were simulated. Figure 28 represents the average structure for each potential step of the ATP hydrolytic cycle coloured by the value of their deviation to a reference state (in this case the average **ATP/ATP** dimer over the last 10 ns of simulation). The first step in this cycle corresponds to the ATP binding to the nucleotide unbound state of the dimer and formation of the ATP bound active dimer (state 1). After the dimerization process one ATP is hydrolysed, originating ADP and IP in one binding pocket (state 2). At this stage either this IP is released (state 4), followed by the hydrolysis of the second ATP with subsequent release of IP (state 5), or the second ATP is hydrolysed (state 3), followed by the simultaneous release of the two IP molecules (state 5). After hydrolysis and exit of the two IPs, the dimer releases ADP and brings the NBDs back to the resting state, starting the cycle all over again.

From the above reported results it is now possible to have a clearer picture of the conformational changes induced during an ATP hydrolytic cycle in the NBD dimer. First of all, we show that the structural changes induced by nucleotide hydrolysis are small and restricted to specific residues or segments. The different nucleotide bound species (ATP, ADP and IP) seem to induce conformational rearrangements in specific regions, especially on residues 11-19 and in the helical sub-domain, particularly in residues 93-124. As can be observed, all the reported residues are

located in external areas of the dimer. Therefore, these results suggest that after hydrolysis the energy harnessed is transduced from the binding pockets to the exterior part of the protein in a complex manner. Moreover, after performing the structural alignment of the MJ0796 dimer with the X-ray structure of a full length importer, we were able to confirm that the region identified as the one presenting the largest conformation rearrangements upon hydrolysis (residues 93-124) is in direct contact with the transmembrane domains, and most probably involved in interdomain communication. The effect of these movements and the atomic details of the NBD:TMD communication mechanism in complete transporters are questions requiring further study.

Secondly, in the IP containing systems, we could observed that, in all replicates, at least one of the binding sites exhibited an outward movement of the loop comprised of residues 38-43. As this loop moves away the IP species becomes more exposed to the solvent, enhancing its probability of diffusing away from the active site.

A third conclusion from this work refers to the function of the helical sub-domain (residues 93-124), which presents the most pronounced rearrangements upon hydrolysis. After performing the structural alignment of the MJ0796 dimer with the B_{12} full length transporter, we observed that the helical sub-domain is in direct contact with the membrane domains in the complete transporter, being involved in NBD:TMD signal transmission and playing an essential role in inter-domain communication.

3.6. Acknowledgements

This work was supported by a FCT (Fundação para a Ciência e Tecnologia) fellowship (SFRH/BD/21433/2005).

4. Conformational changes in the Sav1866 exporter

This work was submitted to the Journal of Physical Chemistry B: Ana S. Oliveira, António M. Baptista, and Cláudio M. Soares, (2010) "**Conformational changes induced by ATP-hydrolysis in an ABC transporter: a molecular dynamics study of the Sav1866 exporter**".

4.1. Abstract

ATP-Binding Cassette (ABC) transporters are ubiquitous membrane proteins that use the energy from ATP binding or/and hydrolysis to actively transport allocrites across lipid membranes against a concentration gradient. The mechanism in the nucleotide binding domains (NBDs) by which ATP binding/hydrolysis powers transport is still unclear, despite all the efforts in the last decade to unravel it. Of special importance is to understand how energy, resulting from ATP hydrolysis, is transmitted from the catalytic to the membrane domains. In this work, we try to identify the ATP-hydrolysis induced conformational changes in a complete ABC exporter (Sav1866) from *Staphylococcus aureaus*, using a molecular dynamics (MD) simulation approach. By performing MD simulations on the ATP and ADP+IP bound states, we identify the conformational consequences of hydrolysis. The major rearrangements upon hydrolysis are not restricted to the NBDs, but extend to the transmembrane domains (TMDs) external regions. The residues mostly affected by hydrolysis are located in the extracellular loops 1 and 3, coupling helices 1 and 2, TMD helix 7 and the helical sub-domain region. Our simulations show, for the first time to our knowledge, NBD dimer opening in the ADP+IP state in contrast with all ATP-bound states, within the context of a complete transporter This opening is a consequence of the dissociation of the ABC signature motif from the nucleotide ribose ring and inorganic phosphate. Additionally, in both studied states, we observe the opening of a gate entrance in the intracellular loop (ICL) region leading to the exposure of the TMDs internal cavity to the cytoplasm. In order to clarify whether this opening was large enough to allow allocrite passage, the transmembrane adiabatic energy profile for doxorubicin was determined. Surprisingly, the energy profile, for both states, is overall downhill from the cytoplasmatic to the extracellular side, and the local energy barriers along the TMDs region are relatively small and easy to be overcome by the allocrite. The major difference between states is related to an energy barrier located in the cytoplasmic gate region in the ATP-bound state, which, upon hydrolysis, becomes severely reduced facilitating allocrite passage. These adiabatic energy profiles

allow us to understand at a molecular level the unidirectionality in allocrite translocation and may give us some new insights on how transport is “powered” in this protein family.

4.2. Introduction

The ATP-binding cassette (ABC) transporters are a family of integral membrane proteins that are able to use the energy from ATP hydrolysis [110, 153] to actively drive substrates (named allocrites) across lipid membranes against the concentration gradient [86]. The allocrites are chemically diverse, ranging from small ions, to sugars, amino acids, lipids, drugs, polypeptides and to polysaccharides [86].

ABC transporters are one of the largest families known and its members are found virtually in all living organisms [72]. The functions of some ABC members are linked to important physiological processes, ranging from the import of nutrients to the export of cytotoxic compounds from cells. Some ABC exporters are known to contribute to multidrug resistance in bacteria, fungi, yeasts, parasites and mammals [15] (particularly to the resistance of tumor cells to anticancer agents [246]). Mutations in several members of the ABC family have been associated with genetic diseases in humans, such as cystic fibrosis [22, 245] and Tangier disease [270], prompting a larger interest in the study of this family of proteins. However, and despite the large amount of experimental and theoretical data available for several ABC family members, the molecular understanding of the unidirectional transport of substrates by ABC transporters is still incomplete.

Despite its cellular function and the diversity of transported allocrites, all the members of this family are formed by a minimal basic functional unit, consisting of a pair of highly hydrophobic trans-membrane modules (TMDs) and two cytoplasmic nucleotide binding modules (NBDs). These four modules can be expressed as single and independent proteins or can be fused to each other in all possible combinations [271]. Usually, in bacterial ABC exporters (such as the Sav1866 exporter), a single gene encodes both the TMDs and the NBD [7, 272, 273]. The TMDs form the permeation pathway for the allocrite [86, 247, 248] and they present high sequence variability between different family members, presumably depending

on the allocrite. The catalytic ATPase modules share high sequence homology (more than 30% identity) and structural conservation between different family members. In the NBDs, several highly conserved sequence motifs, involved in the ATP binding/hydrolysis [86, 247, 248] and in inter-domain communication [268] can be identified. This is the case of the Walker-A motif (also named P-loop) [86, 247, 248], the Walker-B motif [86, 247, 248], the ABC signature motif (LSGGQ) [86, 247, 248], the A-loop (an aromatic residue interacting with the adenine base of the bound nucleotide) [77], two small sequences containing a conserved glutamine (Q-loop) [249] and a histidine (H-loop) [86, 247, 248], the D-loop [7] and the TEVGERV motif (only identified in exporters) [7]. Functional studies of several ABC family members have provided evidence that the two NBDs interact and work as a functional unit [174].

Despite the existence of several high-resolution X-ray data for several full-length transporters, namely for the vitamin B_{12} transporter BtuCD [57, 61], Sav1866 [58, 59], Hl1470/1 [63], $ModB_2C_2$ [60], MsbA [64], molybdate/tungstate transporter [250], methionine transporter [66], P-glycoprotein [65] and the maltose transporter $MalFGK_2E$ [62, 142], many fundamental questions about this family remain unanswered. In particular, it is not clear how the energy released from ATP hydrolysis is harnessed and transferred to the TMD, in order to achieve unidirectional allocrite transport through the membrane. Moreover, the conformational rearrangements induced by hydrolysis are mostly uncharacterized and unknown yet. To clarify these issues and to have a detailed understanding of the transport mechanism, the elucidation of the atomic movements associated with ATP hydrolysis and the transmission of these rearrangements to the TMDs is required. For that purpose, the use of MD simulation techniques (with sufficient simulation time and conformational sampling) appears to be a good alternative to study the dynamic behavior of these transporters. Despite the fact that the timescale for the entire transport cycle is beyond the timescale allowed by present day classic all-atom MD simulations, this technique can be used to study individual

states or initial transitions in the transport cycle. Over the last decade, numerous theoretical studies using MD simulations have been reported for several ABC members, whether using the isolated ATPase domains only, or the full-length transporters. The dynamic behavior and the conformational changes induced during an ATP-hydrolysis cycle in the monomeric and in the dimeric NBDs have been extensively studied over the years [56, 87, 94, 96, 97, 171-173] including by ourselves [1]. More recently, several simulations for complete ABC transporters have also been reported [95, 163, 165, 168-170].

Sav1866 is a bacterial multidrug ABC protein belonging to the Lipid flippase-like transporter family found in bacteria *Staphylococcus aureaus* [58]. The study of this protein is of high clinical relevance since it is a homologue of the multidrug ABC transporters (such as, P-glycoprotein and TAP1/TAP2) [58]. Based on the available Sav1866 crystallographic structures [58, 59], we know that this protein is a homodimer formed by two subunits, each consisting in a unique polypeptide chain fusing the TMD with the NBD (see Figure 29). Each subunit consists of a TMD part (the first 320 amino-acids) and a NBD part (residues from 337 to 578). Moreover, a link region connects the TMD (TM6 helix) to the catalytic NBD. In the presence of ATP molecules, the two NBDs dimerise in a head-to-tail arrangement, with the two nucleotide molecules interacting with the P-loop residues of one monomer and the ABC signature motif of the other (see Figure 29C). As can be seen in Figure 29A and Figure 29B, the TMDs are formed by six trans-membrane spanning α-helices each, yielding a total of 12 helices in the complete transporter (as typical for this family [272]). Additionally, the trans-membrane helices are connected to each other by intracellular (ICL) and extracellular (ECL) loops. The intracellular loops (ICLs) are quite long and extend themselves beyond the lipid bilayer, approximately 25 Å to the cytoplasm [58]. Due to the existence of the ICLs, the NBDs are located far apart from the membrane, and the interface between the NBDs and the TMDs occurs through the intracellular loops (especially ICL_1 and ICL_2) mainly [58].

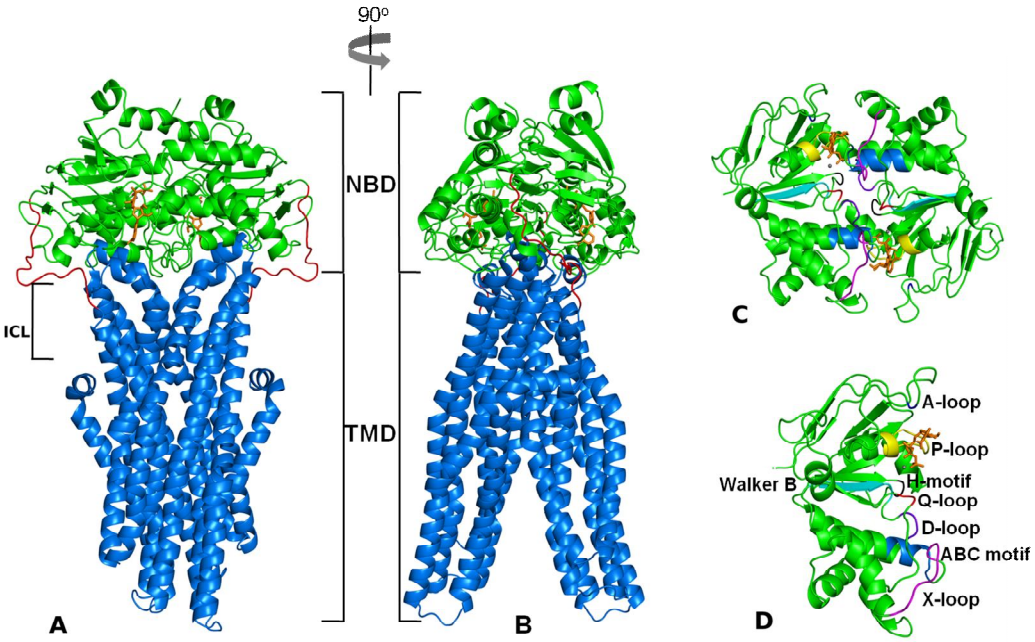


Figure 29- Crystallographic structure of the complete Sav1866 transporter (PDB code: 2HYD) [58]. The NBDs are colored in green, while the TMDs are colored blue. The linker region is highlighted in red. ADP is represented by orange sticks and the magnesium cofactor is a gray sphere. Figures were generated with the program PyMOL [252]. **A**- View of the Sav1866 transporter. **B**- Rotated (90º) view of the Sav1866 transporter. **C**- Structure of the Sav1866 NBD dimer with MgADP bound to catalytic site. The conserved sequence family motifs are colored in red (Q-loop), blue (ABC signature), yellow (Q-loop), magenta (A-loop), cyan (Walker-B), black (H-motif), purple (D-loop) and dark blue (X-loop, which is a motif only observe in exporters). **D**- Detail view of the individual monomer.

Recently, Sav1866 was the subject of a simulation study by Aittoniemi *et al.* [168], where conventional MD simulations were used to study the transient asymmetric transitions in the nucleotide binding sites in a pre-hydrolytic (ATP-bound) state. From multiple 30-40 ns MD simulations, the authors were able to show that an initially symmetrical ATP-bound Sav1866 conformation rapidly evolves to an asymmetrical conformation at the nucleotide binding sites and that these asymmetries affect the TMD-NBD interface. In the present work, we report

extended nanosecond timescale MD simulations of the complete Sav1866 model exporter inserted into a lipid bilayer in a pre-hydrolysis state (ATP-bound) and in a post-hydrolysis state (ADP and inorganic phosphate-bound). With this we wanted to characterise the initial conformational changes that are consequence of ATP hydrolysis, in both the TMDs and the NBDs, allowing us to gain insights into the NBD-TMD communication mechanism. We aim also at comparing the simulated conformational changes observed in isolated NBD dimers [1, 87, 96, 97, 172], with the ones observed in a full-length transporter. Additionally, the permeation properties of the transmembrane channel towards a known substrate, doxorubicin, were studied in the two situations, providing clues on the transport directionality at a molecular level.

4.3. Materials and Methods

4.3.1. Starting Structure

The 3.0 Å resolution crystal structure of the Sav1866 dimer bound with MgADP (PDB code: 2HYD) [58] was used as the starting point for this work. Although this particular structure was crystallized in the presence of ADP, the NBDs exhibit an ATP-bound conformation [58, 59], as was later confirmed by the crystal structure of Sav1866 with a non-hydrolysable ATP analogue [59]. First, we started by recreating the γ-phosphate position, by aligning the NBDs of Sav1866 (crystallized with ADP only) with the ABC transporter MJ0786 NBD dimer [36] (crystallized with ATP bound in the binding sites), and, in this way, determining the missing coordinates. For the post hydrolysis system, with ADP and inorganic phosphate (IP) in the binding sites, the nucleotide coordinates were obtained by eliminating the γ-phosphate coordinates from the tail of ATP, while the positions of the IP atoms were the same as the positions occupied by the corresponding γ-phosphate group of ATP.

The protonation state of each individual protonable group in the protein and in the nucleotide species was determined using a combination of Poisson-Boltzmann calculations, performed with the package MEAD (version2.2.0) [234, 240, 256], and Metropolis Monte Carlo simulations, using the program PETIT (version1.3) [239]. For details related with the determination of the protonation state of the protonable residues, see the Supporting Information section 9.1.

4.3.2. Sav1866 and lipid bilayer assembly

The simulated system consisted of the Sav1866 ATP-bound structure [58] embedded in an equilibrated dimysristoylphosphatidylcholine (DMPC) lipid bilayer (for details related with the membrane construction and equilibration as well as its characterization, see the Supporting Information section 9.2). The TMDs were

inserted in the pre-equilibrated DMPC bilayer and the optimal position of the protein, relative to the membrane, was determined using the location of the charged residues in the transmembrane helices as a reference.

Initially, the TMDs were placed in the membrane by overlaying, both the membrane and the protein, and then all lipids within a cut-off distance of 1.2 Å from the protein atoms were deleted, as described by other authors [274]. The resultant system was then hydrated in a rectangular box, using a pre-equilibrated box of water molecules. The central cavity of the TMDs was also filled with water molecules. The water molecules misplaced in the center of the bilayer (which is formed by the highly hydrophobic DMPC tails), were removed after visual inspection. The final system contains the Sav1866 exporter (with MgATP) embedded in a bilayer formed by 431 DMPC and surrounded by 61726 water molecules, in a total of 216806 atoms. The final box dimensions are 12.9x12.9x15.2 nm.

4.3.3. General setup for the molecular dynamics simulations

All MD simulations were performed using the GROMACS 3.3.1 package [213, 257] and the 53A6 GROMOS96 force field [196, 197, 199, 200]. The parameters for DMPC lipid molecules were taken from the 53A6 GROMOS96 force-field, except for the atomic partial charges, which were the ones derived by Chiu *et al.* [275]. The parameters for the nucleotides species (ATP, ADP and IP) were taken from our previously published work [1]. The single point charge (SPC) model [258] was used for water molecules. Simulations were performed in the isobaric and isothermal ensemble. The temperature of the system was kept around 310K, by coupling to a Berendsen heat bath [261] with a coupling constant of 0.1 ps, with separate coupling of solutes (protein, nucleotides and lipids) and solvent. The chosen temperature for the simulations is well above the phase transition temperature for DMPC (Tm=296-297K) in order to ensure that the bilayer is in the liquid crystalline state [276]. The pressure was coupled semi-isotropically (coupling constant of 1.0

ps and compressibility of $4.5x10^{-5}$ bar^{-1}), resulting in independent coupling of the lateral P(x+y) and perpendicular (Pz) pressures. The values of the lateral and perpendicular pressures were chosen in order to correspond to a surface tension of approximately 25 dynes/cm (which was shown to give the correct area per lipid using the 45A3 GROMOS force-field [277]). The SETTLE algorithm [217] was used for keeping the water molecules bond lengths at their equilibrium values and the LINCS algorithm [216] was employed to keep all remaining bonds constrained. The non-bonded interactions were calculated using a twin range method [192] with short and long-range cut-offs of 8 Å and 14 Å, respectively. The time step for integrating the equations of motion was 0.002 ps. The truncated long-range electrostatic interactions were corrected with a reaction field using a dielectric constant of 54.0 [204, 278].

The ATP-bound system (hereafter and for simplicity purposes named **2ATP**) was first energy minimized to remove excessive strain. We start by performing 5000 steps of steepest descent minimization with harmonic restraints (with a force constant of 1000 kJ mol^{-1} nm^{-2}) applied to all non-hydrogen atoms (including the ones from DMPC lipids), followed by another 5000 steps of the same algorithm, only restraining the protein non-hydrogen atoms, ending with 5000 steps with restraints applied to Cα atoms only. After the minimization procedure, and in order to allow proper repacking of the lipids around the protein, we performed 500 ps of MD simulation with all protein, nucleotides and cofactors atoms harmonically restrained (the force constant used was 1000 kJ mol^{-1} nm^{-2}) at the constant temperature of 310K and constant pressure of 1 atmosphere. Afterwards, 250 ps of MD simulation were calculated, with position restraints (the force constant used was 750 kJ mol^{-1} nm^{-2}) applied to the protein non-hydrogen atoms only. Finally, only the Cα atoms were restrained (the force constant used was 500 kJ mol^{-1} nm^{-2}) for a period of 250 ps. Unrestrained simulations for the **2ATP** state started from this point onward.

After 13 ns of unrestrained MD simulations for the **2ATP** state, a new system was built by transforming the original ATP nucleotides in ADP+IP using a “slow growth” method, as reported elsewhere [98]. In this method, our system hamiltoniam is coupled to a lambda (λ) parameter, which is function of time [190]. Therefore, a continuous change of λ between 0 and 1 allowed the conversion from the ATP to the ADP+IP bound state, in a 5 ps time interval.

In order to reduce the sampling problems in protein simulations, ten 60 ns long MD simulations were calculated for each system in study, in a total of twenty simulations. All replicates were initiated with different sets of random velocities.

4.3.4. Estimation of substrate permeation energy profiles

In order to estimate the energy profile along the substrate permeation pathway, for both the pre- and post- hydrolysis states, we used Monte Carlo (MC) simulations to explore the doxorubicin conformational and positional space inside the Sav1866 transporter, along this profile.

Doxorubicin is an anthracycline drug, which is transported by several exporters (such as Sav1866 and P-glycoprotein), both in the neutral and in the positively charged form [279]. Nevertheless, at physiological pH, around 94% of the doxorubicin molecules are positively charged at the daunosamine group [280] (see Supporting Information section 9.3). By this reason, the minimum of the adiabatic profile was determined for the charged doxorubicin molecule only. The partial charges for the doxorubicin molecule were calculated with the program Gaussian03 [281] and RESP fitting [282], as described in the Supporting Information section 9.3. For the minimum energy profile construction, snapshots were taken from the 60ns MD simulations of replicates 1 to 3, for both **2ATP** and **2ADP.IP** states. Conformations were taken each nanosecond. Due to the large number of conformations to study (366 in total), we performed a first MC screening step in

order to identify the most promising candidates for the MC doxorubicin simulations. For this screening step, a simplified model of the doxorubicin molecule was used, consisting in a non-charged sphere with a van der Waals radius of 5 Å. The program AUTODOCK 4.2.2 [283] was used to perform all MC simulations. The definition of the grid dimensions was critical, since we needed to ensure that the grid covered the complete transmembrane pathway. Each grid set was composed of 50x50x90 points with 0.750 Å spacing. The starting position for the simplified doxorubicin model was the extracellular end of the box. Each MC run consisted in only one cycle with a maximum of 10^6 accepted or rejected Monte Carlo steps at the temperature of 9600K. This temperature was chosen in order to assure that the sphere model would completely traverse the transmembrane pathway enabling an adequate sampling of the system. 50 independent MC runs were performed for each protein conformation. We then use the results of these 50 runs to select the minimum energy conformations along the transmembrane pore axis in order to construct an adiabatic profile for each one of the 366 initial conformations. Based on these adiabatic profiles, only the conformations that allowed the complete sphere passage from the extracellular to the cytoplasmic side of the membrane and, simultaneously, presented energy barriers lower than 15 kcal/mol were chosen as candidates for the doxorubicin adiabatic profile construction. A total of 102 conformations were chosen in this way, corresponding to 43 and 59 for the **2ATP** and the **2ADP.IP** states, respectively.

The MC procedure for the construction of the doxorubicin adiabatic profile, was very similar to one described above, but the grid dimensions were 100x100x180 points with 0.375 Å spacing, and each MC run consisted in only one cycle with a maximum of $7x10^7$ accepted or rejected steps, at the temperature of 2400K. A finer grid size was used in this case, in order to improve the quality of the MC results. The temperature of 2400K was chosen in order to ensure that the doxorubicin molecule was able to overcome the cytoplasmic gate energy barriers, assuring an adequate sampling in the transmembrane pathway. The initial maximum quaternion

rotation step was 5.0 degrees and the initial maximum translational step was 0.2 Angstroms, respectively. Again, 50 independent MC runs were performed for each protein conformation. An adiabatic profile for doxorubicin permeation was built for each nucleotide bound state, and it is characterized by the minimum energy conformations along the z-axis.

4.3.5. Data analysis

The atomic positional deviations were calculated by comparing the **2ADP.IP** with the **2ATP** states within the same replicate, and averaging these differences over all replicates in order to eliminate the structural differences arising from natural variation between replicates (for more details related with this methodology see [1]).

4.4. Results

The main goal of this study is to investigate the conformational reorganization induced by nucleotide hydrolysis, both in the catalytic and in the transmembrane domains, and to clarify the regions involved in inter-domain communication. To this end, a complete ABC exporter (Sav1866) was simulated in two possible states of the ATP-hydrolytic cycle: a pre-hydrolysis state (where two ATP molecules are bound in the binding pockets) and a post-hydrolysis state (with ADP and IP bound to both active pockets). For each state, a total of 10 replicates (each 60 ns long) were calculated to explore the protein dynamics.

For all replicates, the protein behaviour was carefully examined during the simulation by visual inspection and by monitoring several system properties (data not shown), such as the root mean square deviation (RMSD) from the X-ray structure, the number of H-bonds formed between the two monomers and the solvent accessible surface (SASA). In seven replicates, substantial deformations in the tertiary structure were observed (especially in the TMDs), which led to a bending of the protein towards the membrane (see Figure 52 in the Supporting Information section 9.4). At this stage, we cannot clarify the physical reality behind these structural motions, leading to the approximation of the NBD domains towards the membrane. The timescale of the simulations prevents us to infer about reversibility of these motions, given that they are 60 ns long and such phenomena may occur at longer timescales. One thing that is thought provoking is the observation, in some of these cases (results not shown), that the cytoplasmic gate entrance approaches the membrane, hypothetically allowing the transport of allocrites associated with it. However, given that the observed conformational changes are diverse between replicates and nucleotide bound states, we decided to focus our analysis on the replicates that are less deformed, both in the ATP as well as in the ADP+IP simulations (remember that the latter ones are derived from the former). This conservative approach allows us to investigate the conformational consequences of ATP hydrolysis, without having to handle with the “noise” coming

from these bending motions. The simulations in this situation correspond to replicates 1 to 3.

4.4.1. Structural differences between 2ATP and 2ADP.IP state

For the three replicates considered, a simple measure of the overall protein stability can be obtained by plotting the drift from the X-ray structure, using the root mean square deviation (RMSD) of the C_α atoms as a function of time (Figure 30).

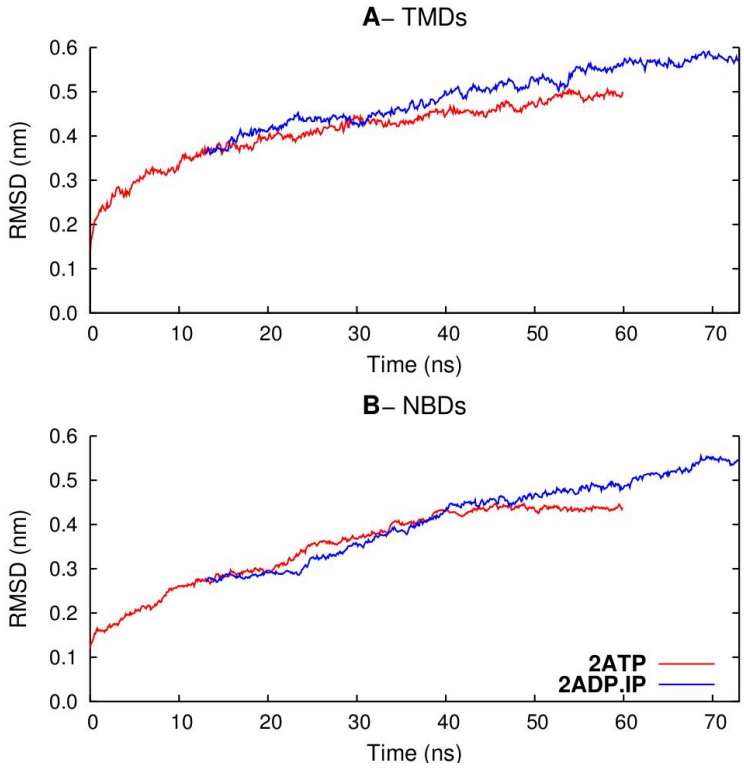


Figure 30- Time evolution of the average Cα RMSD relative to the X-ray structure during the 60 ns of the simulation time for **A**) the TMDs and **B**) the NBDs. The RMSDs were calculated by fitting each domain separately. The RMSD values were averaged from all 3 replicates selected for each state. The red line corresponds to the simulations of the **2ATP** system and the blue line to the **2ADP.IP** system. Each point in the plots represents the average value for 100 ps.

After the removal of the restrains, the RMSD increases continuously, reaching the global values of about 0.49 nm (**2ATP**) and 0.57 nm (**2ADP.IP**) for the TMDs, and 0.44 nm (**2ATP**) and 0.53 nm (**2ADP.IP**) for the NBDs**.** Overall, the average RMSD values observed are slightly higher than the values reported in previous MD studies for ABC exporters [162, 164, 168], but our simulations can be considered stable after 50 ns. Nevertheless, it is important to refer that the lipid environment in which the transmembrane domains are inserted significantly influences the conformations adopted by the protein and, consequently, the RMSD behaviour [168]. In the **2ADP.IP** state, the higher RMSD values observed (compared to the **2ATP** state) reflect the conformational changes that occur upon hydrolysis. These conformational changes will be widely debated in the following sections.

It is currently proposed that nucleotide hydrolysis induces conformational changes in the NBDs, which are, by a still unknown molecular mechanism, transmitted afterwards to the TMDs. Several theoretical works performed in the isolated NBD dimers [1, 87, 97], have already suggested possible NBD:TMD signal transmission events, which need to be validated in the context of a complete transporter. These hydrolysis-induced rearrangements might be at the origin of the unidirectional allocrite transport. Therefore a comparison between the ATP-bound and the ADP.IP-bound states is necessary. For this purpose, the C_α average positional deviations for the **2ADP.IP** system relative to the **2ATP** system were determined as a function of the residue number, for the last 10 ns of simulation. This is a simple and straightforward way to identify the residues that present high difference between states. The results can be observed in Figure 31. In this plot, the values reported were determined by comparing the **2ADP.IP** average structure for each replicate to the **2ATP** average for the same replicate (see the materials and methods section for details).

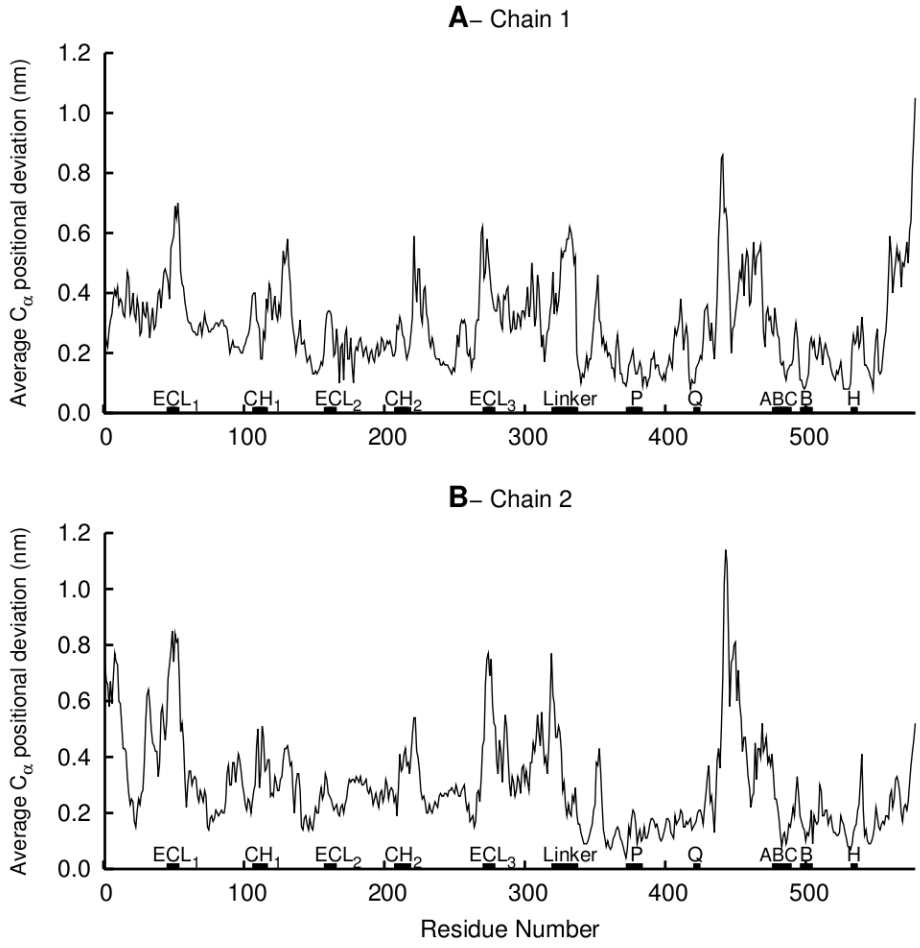


Figure 31- Average Cα positional deviation (over the last 10 ns of the simulations) with respect to the average **2ATP** conformation, obtained by comparison within replicates. The values were obtained by comparing the average structure for each **2ADP.IP** replicate with the **2ATP** average structure for the same replicate. Then, these individual deviations were averaged over all three replicates. Each chain consists of a transmembrane (residues 1-320) and a catalytic (337-578) domain. These two domains are connected by a linker region (residues 321-336). The structural family motifs located in the catalytic domains are marked in the plots, namely the P-loop (P), Q-loop (Q), ABC signature motif (ABC), Walker-B motif (B) and the H-motif (H). The region connecting the TMDs to the NBDs is identified as "Linker". For the TMDs, the coupling helices 1 and 2 and the extracellular loops 1, 2 and 3 are marked with the acronyms CH_1, CH_2, ECL_1, ECL_2 and ECL_3, respectively.

As can be observed from the comparison of Figure 31A and Figure 31B, the behaviour of the two monomers is more or less symmetric, with the regions presenting higher average Cα deviation values being the same in the two chains. There are however some regions, such as the C-terminus of Chain 1 and the N-terminus of Chain 2, that are exceptions, and are different between chains. We cannot determine if this localized asymmetry arises from the X-ray structure asymmetry or from the different dynamic behaviour of the two chains during the simulation.

The regions presenting the highest average C_α deviation values between the ATP-bound and the ADP.IP-bound states (average positional deviation ≥ 0.50 nm) are the residues in segments 41-56, 221-222, 269-279, 318-336, 438-454 and 464-468. Additionally, for Chain 1 (Figure 31A), the 128-131 segment and the C-terminus region also present high average C_α deviation values. For Chain 2 (Figure 31B), the segment 110-113 and the N-terminus region have also high positional differences between states. Mapping these differences in the average three-dimensional structure for the **2ADP.IP** system, we can identify the spatial location of the residues that present the largest displacements during the ATP to ADP+IP transition (Figure 32).

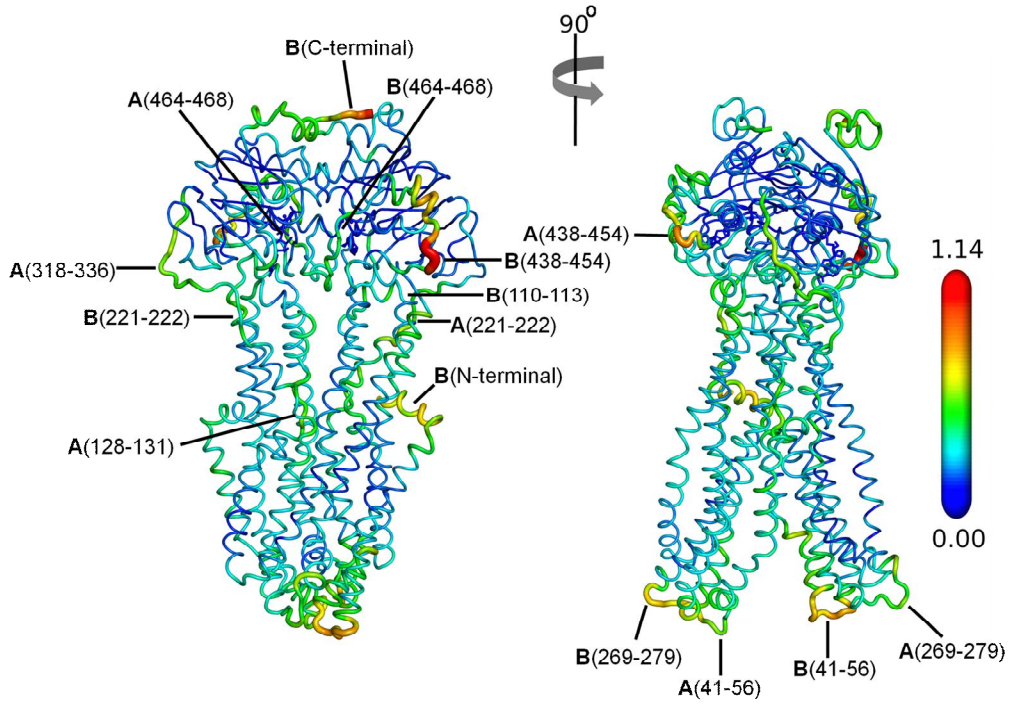


Figure 32- Mapping of the Cα positional deviation values (relative to the **2ATP** state), plotted in Figure 31, on the average **2ADP.IP** structure. The average structure was determined using the last 10 ns of simulations of the three **2ADP.IP** replicates, after fitting to the X-ray structure. Figures were generated with the program PyMOL [252]. The average **2ADP.IP** structure (Chain 1 and 2) is coloured according to the Cα positional deviation gradient (in nm). The regions with higher deviations are colored in red while the ones with lower deviations are colored in dark blue. The protein is rendered as a main chain cartoon; the ADP and IP are represented with sticks while the spheres correspond to the Mg ions. The regions presenting high deviations relative to the **2ATP** state are represented with a higher main chain radius, while the ones with lower deviation are thinner. In this figure, the letters A and B in the legends, refer to chain A and B respectively.

The set of residues 41-56, 110-113, 128-131, 221-222 and 269-279 are located in the TMD region of the protein, while residues 318-336 are located in the linker region (which is a small polypeptide loop that directly connects the TMDs to the NBDs). Residues 41-56 are part of the extracellular loop 1 (ECL_1), while residues

269-279 form the extracellular loop 3 (ECL_3), and these two ECLs are located at the end of the TMDs helices in the cytoplasmic side of the membrane. The aminoacids ranging from 110-113 are located in the coupling helix 1 (CH_1) [58] while residues 221-222 (in transmembrane helix 12) are the residues contiguous to the coupling helix 2 (CH_2). The coupling helices are small alpha helices oriented roughly in parallel to the lipid bilayer and that form the TMD:NBD interface. These helices were suggested to be involved in inter-domain conformational transmission due to its direct contact with the catalytic domains [58] The 128-131 segment is positioned in the helix 7, which is located in the narrowest part of the transmembrane pathway. In the X-ray structure, the residues from both chains are closing the transmembrane channel from the cytoplasmatic side. The 438-454 and 463-468 segments, as well as the C-terminus are situated in the NBD region of the protein. Residues 438-454 and 463-468 are located in the helical sub-domain (HD) region, which is in direct contact with the TMD coupling helix 2, and has been suggested to be involved in inter-domain communication [1, 87, 173, 267]. Moreover, the 463-468 segment forms a structural arrangement, called X-loop, which has been suggested to be of high functional importance in inter-domain communication [58, 268].

From Figure 32, we can observe that the HD regions present the most pronounced conformational differences between the **2ATP** and the **2ADP.IP** states. This observation is in agreement with previous simulation works carried out in the isolated catalytic domains of other ABC members [1, 87, 264], where the helical sub-domain was evidenced as one of the regions presenting the largest conformational transitions upon nucleotide hydrolysis.

One interesting observation is the fact that the major conformational changes upon hydrolysis are not restricted to the areas surrounding the binding site or the catalytic domains. As can be seen in Figure 31 and Figure 32, the nucleotide binding family motifs do not present a high positional deviation value in the **2ADP.IP** state, relative to the **2ATP** system. Additionally, we were able to identify several conformational

rearrangements in the TMDs, for example in the ECL (1 and 3) region, which are located more than 50 Å away from the nucleotide binding site (where the actual hydrolysis occurs). It is now clear that the effect of nucleotide hydrolysis can be easily transmitted to the rest of the transporter.

4.4.2. Opening of the NBDs dimer upon hydrolysis

Nowadays, it is generally accepted that ATP binding is essential for the formation of the closed NBD dimer, with the two ATP molecules functioning as a molecular "glue", keeping the catalytic monomers close together [35, 56, 284]. Moreover, it has been suggested that, after hydrolysis, the monomers move apart and the post-hydrolysis products are released. Although, in the last years, an atomic picture supporting NBD dimer opening started to emerge [96, 97], the details of this opening in the context of a complete transporter are not yet known. Wen and Tajkhorshid [96], in their work with the Maltose importer NBD dimer (MalK), were able to observe an opening of the dimer interface upon ATP hydrolysis (regardless hydrolysis occurring in one or the two binding sites), in a time scale ranging from 30 to 50ns. Jones and George [97] reported an opening of the ADP-bound binding sites for the MJ0796 lipid transporter in a timescale ranging from 8 to 28 ns. However, for the same MJ0796 system, we could not observed a similar NBD opening event in the 30 ns timescale, but at a lower temperature [1]. Nevertheless, in complete transporters, where the NBDs are coupled to the TMDs, the NBD motions can be slower, and the dimer opening mechanism can be significantly different from the ones previously observed.

In our simulations, and in order to clarify whether we could observe the NBD dimer opening within the simulated timescale, the contact area between the NBD monomers was determined, for the **2ATP** and the **2ADP.IP** states (results not shown). For the **2ATP** state the contact area remained more or less constant over time, which indicates a highly stable dimer interface. As reported in literature [35,

56, 96, 284], it seems that the ATP molecule stabilizes the NBDs in a close conformation, through a network of interactions between the two monomers. In contrast, for the **2ADP.IP** state, one of the replicates (replicate 3) presented a significative decrease in the contact area, mainly due to the opening of site 2 (see Figure 33 for an illustrative example).

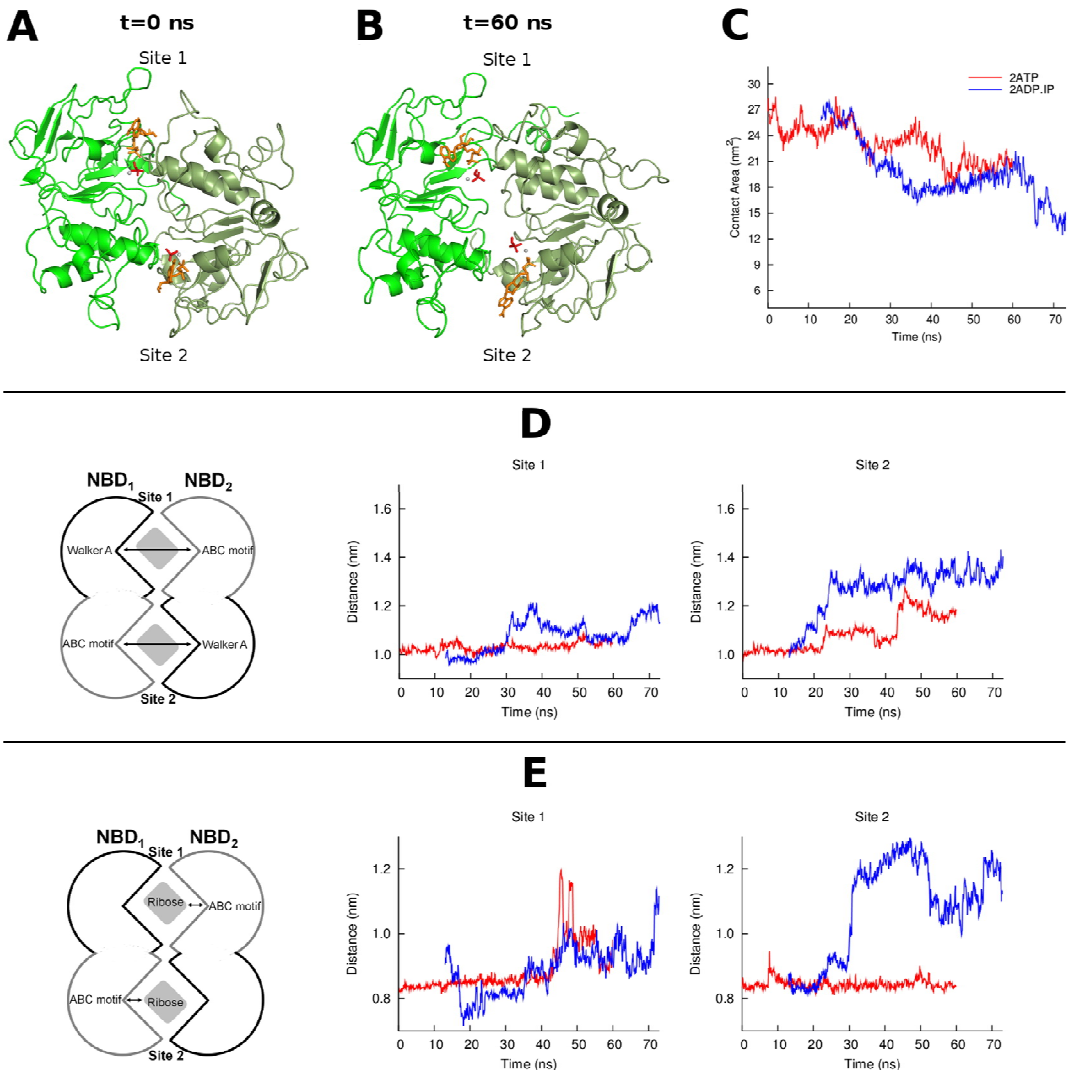


Figure 33- Opening of the NBD interface for one **2ADP.IP** replicate (replicate 3). **A**- Structure for the initial (t=0 ns) NBD dimer. **B**- Structure for the final (t=60ns) NBD dimer. NBD_1 is coloured in light green while NBD_2 is in dark green. ADP and IP are represented by orange and red sticks, respectively. The gray spheres represents the Mg ions. **C**- Time evolution of the contact area between NBD_1 and NBD_2 of the same replicate, for the **2ATP** and the **2ADP.IP** states. The nucleotides, IPs and magnesium ions where used in the SAS calculations. **D**- Distance between the Walker-A motif of one NBD and the ABC signature motif of the other. **E**- Distance between the nucleotide ribose rings and the ABC signature motif residues. In figures **C**, **D** and **E**, the red line corresponds to the **2ATP** state while the blue line represents the **2ADP.IP** state. Each point in the plots represents the average value over 100 ps.

As can be seen in Figure 33A, in the beginning of the **2ADP.IP** simulation (replicate 3), both binding sites are closed, and the post-hydrolysis products are sandwiched between the P-loop and the ABC residues. However, after 60 ns of simulation (Figure 33B), site 2 exhibits a clear interface opening and dissociation of the ABC signature motif from ADP. Nevertheless, in the end of the simulation, both post-hydrolysis products remain associated to the corresponding P-loop. This can be quantitatively illustrated in Figure 33C, where we can observe that the NBD interface reduction starts 8 ns after the replacement of ATP by ADP+IP.

Additionally, we also determined the contact area between the NBD monomers in the seven replicates that presented bending (data not shown). For the **2ATP** state, the contact area remained constant over time and no dimer dissociation was observable, whereas in two replicates of the **2ADP.IP** state, a clear reduction in the contact area was present, mainly due to the NBD interface opening. In these two **2ADP.IP** replicates, and similarly to the described example on Figure 33, dissociation of the ABC signature motif from ADP is only present in binding site 2.

To quantify the NBD dimer separation in replicate 3, the distance between Walker-A and the ABC signature motif of the opposing monomer was measured, both for the **2ATP** and the **2ADP.IP** states (Figure 33D). From this figure, there is no doubt that

dimer interface dissociation only occurs in site 2 in the **ADP.IP** state, which presents an increase in the above described distance, when compared to the **2ATP** state. In site 1, whether for the pre- and post- hydrolysis states, no relevant change in the dimer interface was observed. This is interesting, since hydrolysis occurred in both binding sites, but only one presents opening. This asymmetric behavior of the NBDs has also been observed in previous theoretical works [56, 96, 97]. Additionally, the distance between the nucleotide ribose ring and the residues that form the ABC signature motif was also determined (Figure 33E). As can be seen, site 2 opening is accompanied by the dissociation of the ribose ring (and the IP species) from the ABC signature residues. These post hydrolysis products dissociate from the ABC motif, but not from the corresponding P-loop (which prevents ADP and IP diffusion away from the binding site). Similar observations were reported by Wen and Tajkhorshid for the MalK dimer opening [96].

A key role for the Q-loop glutamine in hydrolysis is frequently suggested in the literature [36, 51, 285]. The side-chain of this conserved glutamine residue coordinates the magnesium ion located in the binding site, and it may orient the nucleophilic water needed for ATP hydrolysis [51]. Additionally, it is also suggested that the disengagement of the canonical interaction between the Q-loop glutamine and the magnesium ion may be coupled with the HD rotation [51, 87]. However, this coupling is still far from consensual and both structural and computational data available are still inconclusive [1]. In our simulations, both for ATP-bound and ADP.IP-bound states, the evolution of the canonical Q-loop glutamine-Mg interaction, as well as the HD rotation was monitored, hoping to clarify whether there is any relation between these two events. For the **2ATP** state, the canonical Q-loop glutamine-Mg interaction is retained during the 60 ns of simulation, in all three replicates, while it is lost in one **2ADP.IP** replicate (results not shown). The HD region presents a small rotation (especially in monomer1) in all **2ADP.IP** replicates, when compared with the **2ATP** replicates. Additionally, we also monitored the evolution of the canonical Q-loop glutamine-Mg interaction in the

seven replicates where bending is observed (results not shown). We were able to observe that in the **2ATP** state, in all seven replicates, the canonical interaction was kept intact, whereas in the **2ADP.IP** state three replicates presented a withdrawn of the glutamine side chain and the disengagement of this canonical interaction. Nevertheless, there seems to be no relation between the disengagement of the canonical Q-loop-Mg interaction and the movement of the HD region (see Figure 53 of the Supporting Information section 9.5). Additionally, we were unable to identify any relation between the NBD interface dissociation with the Q-loop-Mg withdrawn for the **2ADP.IP** state (see Figure 54 of the Supporting Information section 9.6).

4.4.3. An entry point for the allocrite molecules?

In our simulations of both studied states (although with a higher frequency in the **2ADP.IP** state), we could visually observe the spontaneous opening of a tunnel entrance in the ICL region, which directly connects the cytoplasm side of the membrane to the interior of the TMDs (see Figure 34).

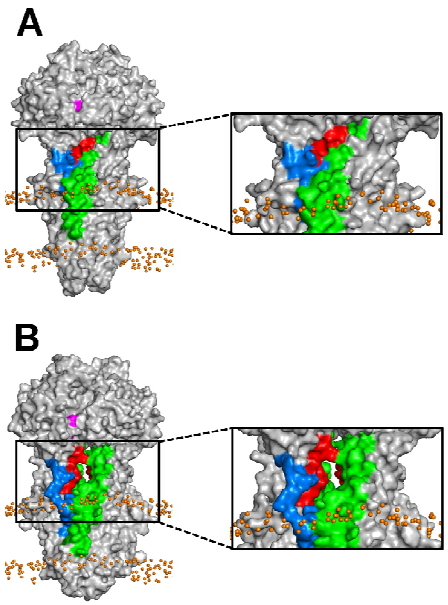


Figure 34- Overview of the cytoplasmic gate entrance to the TMD internal cavity. This figure represents an example of the opening occurring in the **2ADP.IP** replicate 2. Similar movements can be observed for the other replicates. The opening directly connects the cytoplasmic side to the transmembrane pore. This gate is located in the intracellular loop (ICL) region, just below the NBDs. The transmembrane helices 7, 11, 15 are coloured in red, green and blue, respectively, wheras the ADP molecules are highlighted in magenta. **A**- Sav1866 X-ray structure. **B**- Snapshot of the cytoplasmic gate opening for the **2ADPIP** replicate 2 after 24 ns of MD simulation

The residues of the TMDs involved in the gate opening are in direct contact with the X-loops (in the NBDs). This region contains a complex interaction network formed by the side chains of residues from TMD helix 7, 11 and 15 and the side chains of the NBDs X-loops (especially glutamate 473). This interaction network closes the internal TMD cavity from the cytoplasmic side. As helices 7, 11 and 15 move away from their initial position and the cytoplasmic gate entrance opens the TMD internal cavity, making it accessible from the cytoplasmic side (see Figure 34B). Despite the observed movement of the TMD helices 7, 11 and 15, the overall transporter tertiary structure is not altered, and the interaction network between the NBDs X-loop and TMD helix 7 is kept stable through the simulation time.

At this point it seems relevant to refer a recent work by Weng an co-workers in the MsbA exporter [170]. These authors were able to observe the atomic details of the transition between an outward to an inward facing conformation, by using target molecular dynamics simulations. In their case, the cytoplamic gate opening (also located in the ICL region) was a consequence of the dissociation of the NBD dimer interface (the first step in the transition from an out- to an inward conformation). In our work, and although a similar cytoplamic gate entrance was observed, we could not identify any relation with the NBD dimer dissociation. In our case, the cytoplasmic entrance was also identified in the ATP-bound state (where no dissociation of the NBD interface exists). Nevertheless, it is quite curious that the cytoplamic gate entrance observed in our simulations was apparently large enough

to accommodate and allow the passage of some of the known Sav1866 allocrites (such as doxorubicin). This observation leads us to question whether this entrance could have a biological function, for example, in allocrite uptake. In order to clarify this, the adiabatic energy profile for the charged doxorubicin along the transmembrane pathway was computed for several conformations, both for the **2ATP** and the **2ADP.IP** states (see Figure 35).

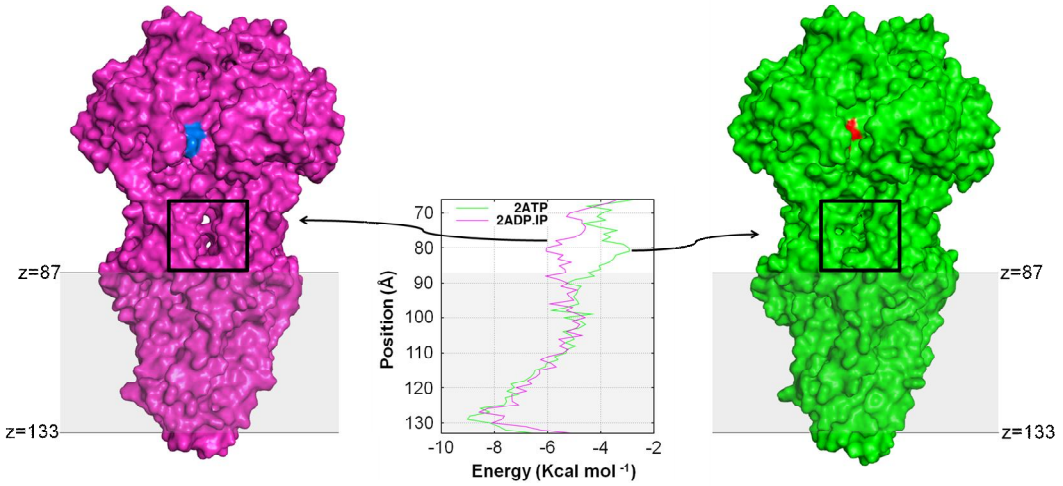


Figure 35- Overall minimum energy profile (kcal/mol) along the transmembrane pore, computed for the **2ATP** (in green) and **2ADP.IP** (in magenta) states. Additionally, the global average structures (obtained over replicate 1, 2 and 3) over the last 50 ns of simulation are also represented for **2ATP** (right) and **2ADP.IP** (left), in order to allow a clear visualization of the cytoplasmic gate opening. The ATP and the ADP nucleotides are coloured in red and blue, respectively.

As can be seen in Figure 35, the energy profile for both states is quite similar in the transmembrane region (coloured in gray). In this region, the energy profile from the part closest to the cellular interior (top part) towards the part closest to the cellular exterior is overall downhill, which fits the notion that the Sav1866 transporter is an exporter. Additionally, this downhill energy profile helps explain the unidirectional

transport in ABC transporters. However, in the ICLs region (where the cytoplasmic gate is located; top part of the profile in Figure 35) the energy profile has significant differences between the ATP-bound and the ADP.IP-bound state. The doxorubicin energy barrier for the pre-hydrolysis state is much higher than for the pos-hydrolysis state, which indicates that doxorubicin more easily crosses the cytoplasmic gate in the **2ADP.IP** state than in the **2ATP** state.

Upon comparison of the global average structures (averaged over replicate 1,2 and 3) for both states (**2ATP** in green and **2ADP.IP** in magenta), we can observe that the cytoplasmic gate entrance for the **2ADP.IP** state is wider, and the transmembrane pore is much more exposed than for the **2ATP** state. For the **2ATP** state, the cytoplasmic gate, when open, is smaller and narrow, which account for the higher energy barrier seen in the adiabatic energy profile.

The observed higher energy barrier for the passage of doxorubicin in the ATP bound state (which can be similar for other substrates), when compared with the ADP.IP-bound state, clearly fits a controlled transport mechanism mediated by ATP hydrolysis, since the transporter opens its cytoplasmic gate upon hydrolysis, allowing the downhill passage of the allocrite. In contrast, ATP binding closes the gate, preventing passage.

4.5. Concluding Remarks

In this work, we have performed extensive MD simulations on the full-length ABC exporter Sav1866 in two distinct states of the ATP cycle: a pre-hydrolysis state (where two ATP molecules are bound to the binding pockets) and a post-hydrolysis state (with ADP and IP bound). For each state, several replicates were performed in order to reduce the sampling problems.

In our simulations, ATP hydrolysis induces structural changes, some of them located in the NBDs. These changes are similar to the ones observed before for isolated NBD dimers in solution [1, 87, 97]. However, the hydrolysis induced structural rearrangements are not restricted to the NBDs and it was possible to identify major conformational changes in the TMDs and in the linker regions. Even the farther regions of the TMDs (such as the extracellular loops) were influenced by the nucleotide state in the binding sites. All these observations reinforce the idea that, in ABC transporters, energy transduction from the binding pockets to the rest of the protein is very complex.

We could also observe the opening of the NBD dimer interface in some of the ADP.IP-bound replicates, while for the ATP-bound state this was not observable. For the **2ADP.IP** replicates where the NBD dissociation occurred, the NBD dimer opening was only identified in one of the binding sites (binding site 2), and was accompanied by a dissociation of the γ-phosphate and the ribose from the ABC signature motif, although these stayed bound to the corresponding P-loop and did not diffuse away from the binding site.

We were also able to observe a cytoplasmic gate entrance opening in the ICL region, which led to the exposure of TMD cavity, making it accessible from the cytoplasmic side. In some snapshots, this opening was large enough to accommodate some of Sav1866 allocrites (like doxorubicin). This was confirmed by the adiabatic energy profile of doxorubicin along the transmembrane pathway which

showed that, in both nucleotide bound states, the interaction energy goes downhill overall, from the cytoplasmatic to the extracellular side, explaining the exporter nature of Sav1866, and the unidirectionality of this transport. The only allocrite controlling zone seen in the energy profile is at the cytoplasmatic gate entrance, which shows a barrier in the ATP-bound state, preventing efficient passage of doxorubicin. This energy barrier disappears in the ADP.IP-bound state, allowing allocrite permeation. This can be the molecular basis for the mechanism of allocrite transport in ABC transporters.

4.6. Acknowledgements

We want to thank Dr. Carlos Cunha for all the help in the CE/MC calculations. This work was supported by a FCT (Fundação para a Ciência e Tecnologia) fellowship (SFRH/BD/21433/2005).

5. Conformational changes in the MalFGK$_2$E importer

This work is presented in the manuscript: Ana S. Oliveira, António M. Baptista, and Cláudio M. Soares, (2010) "**Inter-domain communication mechanisms in an ABC importer: A molecular dynamics study of the MalFGK$_2$E complex**"

5.1. Abstract

ATP-Binding Cassette (ABC) transporters are ubiquitous membrane proteins able to convert the energy from ATP binding or/and hydrolysis into transmembrane movement of substrates against the concentration gradient. Despite the large amount of structural and biochemical data available for the ABC transporter family, the mechanism by which energy obtained from ATP hydrolysis, in the ATPase domains, is "transmitted" to the transmembrane domains, is still unclear. In this work, we focus our attention on the consequences of ATP hydrolysis and inorganic phosphate exit in the maltose uptake system ($MalFGK_2E$) from *E. coli*. The prime goal is to identify and map the structural changes occurring during an ATP-cycle. For that purpose, we use molecular dynamics simulations to study three potential intermediate states for the ATP-hydrolytic cycle: an ATP-bound, an ADP plus inorganic phosphate-bound and an ADP-bound state. Our results show that the residues presenting major rearrangements during the ATP-cycle are located in the A-loop and in the helical sub-domain region of the NBDs, and in the "EAA motif" region (especially in the "coupling helices") of the TMDs. Additionally, we were able to clarify the atomic details of the NBD dimer interface dissociation, in the ADP-bound state, and we show that the interface opening is accompanied by the dissociation of ADP from the ABC signature motif, but not from its corresponding P-loop motif. This work, together with several other MD studies, suggest a general communication mechanism, both for importers and exporters, in which ATP-hydrolysis induces conformational changes in the helical sub-domain region, that are then transferred to the transmembrane domains via the "coupling helices".

5.2. Introduction

The ATP-binding cassette (ABC) transporters superfamily [5, 7, 71] is one of the largest class of transporters known and they are expressed ubiquitously in all kingdoms of life [5, 7, 71]. The members of this family play essential roles in many cellular processes, as they couple the energy gained from ATP hydrolysis [110, 153] to the transport of an enormous variety of solutes, or allocrites, (such as, inorganic ions, peptides, lipids, antibiotics, pharmacological drugs) across cellular membranes against the concentration gradient. Human ABC members are involved in several genetic diseases (such as bleeding disorders [18], eye [19] and liver diseases [21], and cystic fibrosis [22, 245]). There are several ABC members in humans that, when over expressed, confer resistance to a wide variety of chemotherapeutic drugs [143]. Additionally, several ABC transporters have also been implicated in antibiotic resistance in bacteria [143], drug resistance in fungi [143] and herbicide resistance in plants [143].

ABC transporters comprise importers, which translocate allocrites to the cellular interior, and exporters, which do the opposite. Until now, importers have only been found in prokaryotes, whereas exporters are ubiquitously expressed in all kingdoms [14]. Independently of the transport directionality, ABC transporters are usually composed by a minimum “functional core” formed by four modules (Figure 36A): two transmembrane domains (TMDs) and two catalytic domains (NBDs). All four basic domains can be expressed as separated polypeptides or can be fused together, in virtually in all possible combinations [271].

The ABC molecular architecture is nowadays well established (manly due to X-ray crystallography contributions) and it is generally accepted that the TMDs recognize and provide the passageway for allocrites across the membrane, whereas the NBDs bind/hydrolyze ATP and power the transport. The low sequence conservation of the TMDs is thought to reflect the large diversity of allocrites transported, while the NBDs high sequence and structure conservation may suggest a conserved

powering mechanism for all ABC members (both importers and exporters). The NBDs are formed by two distinct sub-domains (the RecA-like and the helical sub-domains) and present several characteristic sequence motifs (Figure 36B) associated with nucleotide binding and hydrolysis (e.g. the P-loop, ABC signature motif, A-loop, H-loop and Walker B motifs) [247, 248]. The two NBDs form a head-to-tail dimer (Figure 36A and Figure 36B) with the two nucleotide molecules bound at the interface between the P-loop of one monomer and the ABC signature motif of the other [51]. ABC importers, additionally to the four basic domains, require a supplementary substrate binding protein, whose function is to capture the allocrite molecules and to deliver them to the transmembrane face of the transporter [7].

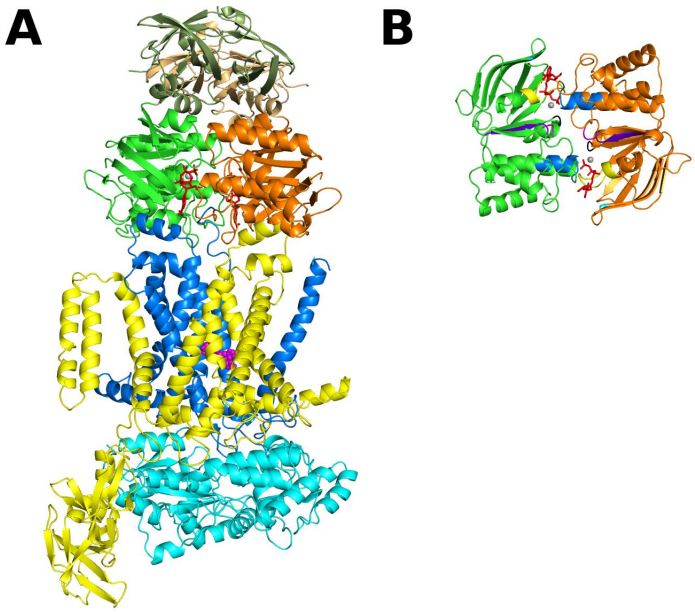


Figure 36- Crystallographic structure for the ATP-bound $MalFGK_2E$ complex (PDB code: 2R6G) [62]. The NBD_1 and the NBD_2 (named $MalK_1$ and $MalK_2$) are colored in green and orange respectively, while the TMD_1 (called MalF) is yellow, the TMD_2 (named MalG) is blue and the MBP (called MalE) is cyan. ATP is represented by red sticks, the magnesium cofactors as gray spheres and the maltose molecule is represented by magenta spheres. All figures were generated with the program PyMOL [252]. **A**- View of the $MalFGK_2E$ complex. In this image, the regulatory domains are colored in lighter green and

orange. **B**- Rotated (90º) view of the MalK dimer. The conserved sequence family motifs are colored: Q-loop (magenta), ABC signature (blue), P-loop (yellow), A-loop (cyan), Walker-B (purple) and H-motif (black). For clarity purposes, the regulatory domains are not represented in this image.

Based on the comparison of several full-length X-ray structures for both importers [57, 60-63, 66, 142, 250] and exporters [58, 59, 64, 65], it is now clear that the structures for these two types of ABC transporters differ significantly, mainly in the transmembrane region. Nevertheless a general transport mechanism has been hypothesized [14] based on the NBDs similarity throughout the family and in the existence of "coupling helices" in the TMDs (small α-helices oriented parallel to the membrane which directly interact with the NBDs), both in importers and exporters. In this unified mechanism, the ATP energy (whether from binding and/or hydrolysis) is converted into conformational changes, which are then transmitted from the NBDs to the TMDs, ultimately allowing active transport of the allocrites [59, 286].

The maltose/maltodextrin uptake system from *E. coli* is one the best functionally characterized ABC member, and transports maltooligosaccharides, up to seven glucose units long [181]. This import system is formed by two different integral membrane proteins (MalF and MalG), two copies of an ABC module (MalK) and a periplasmatic binding protein (MalE) [62] (see Figure 36A). In this transporter, the ABC monomers (MalK) differ from the majority of other ABC ATPases, since they contain an extra regulatory domain (RD) in the C-terminus region, additionally to the highly conserved NBD (see Figure 36A). The RD, which is about 135 residues long, is known to interact with regulatory proteins, such as the transcriptional regulator MalT [287] or the enzyme IIA (a glucose-specific permease from the glucose-phosphotransferase system [288-290]). The RDs play a key role in the stabilization of the NBD dimer and based on the X-ray structures available for the MalK dimer [50, 52, 53, 62, 142] (from different organisms and crystallized in several nucleotide-bound conditions) it is now possible to have a detailed view of the NBD dimer functioning mechanism, and to understand the molecular basis of the

increased stability of the nucleotide-free and nucleotide-bound MalK dimer. This increased stability is associated with the subunit-subunit interactions involving the RDs of both monomers [52]. In the nucleotide-free MalK dimer structure [52, 142], the NBD regions are separated (to different degrees depending whether they are in the open or semi-open state) and the dimer is solely maintained by contacts between the C-terminal RDs. Upon ATP binding [52, 62], the helical sub-domains move forward and close the NBD dimer (with the ATP molecules trapped in the binding sites).

Despite the large amount of experimental data available for the ABC transporter family (including high-resolution X-ray structures for several full-length members [57-66, 142, 250]), many fundamental questions are still waiting to be answered. In particular, how is the energy released from nucleotide hydrolysis converted into mechanical work, in order to allow unidirectional allocrite transport. Another question is how are the conformational rearrangements induced by ATP hydrolysis transmitted from the NBDs to the TMDs, allowing allocrite passage. Understanding the detailed mechanism of transport requires the knowledge of the transporter dynamics and associated conformational changes. Molecular dynamics (MD) simulation techniques (with sufficient simulation time and conformational sampling) are a good way to study the dynamic behavior of the transporters. In the last decade, several studies using MD simulations have been reported (including by ourselves [1]) for several isolated ABC domains [1, 56, 87, 94, 96-98, 171-173] or for complete ABC members [2, 93, 95, 162-170], aiming, not only to study the dynamic behavior of the ABC proteins, but also the structural transitional pathways between conformations. Some of these works even used the isolated ATPase domains of the maltose transporter as a model [94, 96].

The main objective of this work is to map the short time scale response (<50 ns) of an ABC importer to nucleotide hydrolysis and inorganic phosphate exit. For this purpose, we simulated the $MalFGK_2E$ complex in three intermediate states of the

ATP-catalytic cycle: a pre-hydrolysis state (with two ATP molecules bound in the nucleotide binding sites), a post-hydrolysis state (with two ADPs and two inorganic phosphate species bound) and a post-IP exit state (with two ADPs bound). In order to reduce the known sampling problem in MD simulation of proteins [253], ten, 50 ns long, replicates were performed for each system. Based on our simulations, and using the maltose uptake system as a prototype for importers, we were able to identify the ATP-dependent conformational changes, as well as the residues responsible for inter-domain communication in ABC importers. Additionally, we compared the major conformational changes observed for the maltose importer with our previous study on an ABC exporter [2], in order to infer about a potentially general communication mechanism in the ABC transporter family.

5.3. Materials and Methods

5.3.1. Starting Structure

The 2.8 Å resolution crystal structure of the E159Q mutant MalFGK$_2$E transporter (PDB code: 2R6G) [62] was used as the starting point for this work. In this crystal structure, the “hydrolytic” glutamate (located in the nucleotide binding site region, and in close proximity to the γ-phosphate of ATP), was mutated to a glutamine, producing a hydrolysis deficient transporter [62]. In our simulations, the mutation was reverted and the glutamine residue was substituted back to glutamate, in order to recreate the wild type transporter. Additionally, in the X-ray structure, there were several missing atoms and residues, which were modeled with the program MODELLER 6v1 [291] in order to reconstruct the complete transporter structure. For this reconstruction, 10 different models were generated and the one with the lowest value for the MODELLER’s objective function was selected.

The protonation state of each protonable group was determined using a combination of Poisson-Boltzmann calculations, performed with the package MEAD (version 2.2.5) [234, 240, 256], and Metropolis Monte Carlo simulations, using the program PETIT (version1.3) [239]. For details related with the determination of the protonation state of the protonable residues, see the Supporting Information section 10.1.

5.3.2. MalFGK$_2$E insertion into a lipid bilayer

Our reconstructed ATP-bound MalFGK$_2$E complex was inserted in a pre-equilibrated dimysristoylphosphatidylcholine (DMPC) lipid bilayer (for details related with the membrane construction and equilibration as well as its characterization, see the Supporting Information section 10.2). The MalF and MalG optimal position

relative to the membrane was determined based on the hydrophobicity of the protein's residues. After MalF and MalG insertion into the membrane, all lipids within a cut-off distance of 1.2 Å from the protein atoms were rejected, as described by other authors [274]. Subsequently, the system (protein and membrane) was hydrated using a pre-equilibrated box of SPC water molecules [258] in a rectangular box. The central cavity formed by MalF and MalG was also filled with water molecules. The water molecules misplaced in the center of the bilayer (which is formed by the highly hydrophobic DMPC tails), were removed after visual inspection. In the final system, the ATP-bound $MalFGK_2E$ complex is embedded in a 403 DMPC bilayer and surrounded by 77018 water molecules, in a total of 268394 atoms.

5.3.3. General setup for the molecular dynamics simulations

All MD simulations were performed using the GROMACS 3.3.1 package [213, 257] and the 53A6 GROMOS96 force field [196, 197, 199, 200]. The parameters for the nucleotides species (ATP, ADP and IP) were taken from our previously published work [1], whereas the parameters for the maltose molecule were obtain from the 45A4 GROMOS96 force field [292]. The parameters for DMPC lipid molecules were taken from the 53A6 GROMOS96 force-field, except for the atomic partial charges, which were the ones derived by Chiu et al. [275]. All simulations were performed at the constant temperature of 310K. The temperature of the system was coupled using a Berendsen heat bath [261] with a coupling constant of 0.1 ps, and separate coupling of solutes (protein, nucleotides and lipids) and solvent. The chosen temperature for the simulations (310K) is above the phase transition temperature for the DMPC lipids (Tm=296-297K) in order to ensure that the bilayer is in the liquid crystalline state [276]. The pressure was coupled semi-isotropically (coupling constant of 6.0 ps and compressibility of 4.5×10^{-5} bar^{-1}), resulting in independent coupling of the lateral P(x+y) and perpendicular (Pz) pressures. For all simulations, the z pressure component was kept at 1 atm, and the x and y components were

calculated in order to obtain a surface tension of 25 dynes/cm (which was shown to give the correct area for the DPPC lipid using the 45A3 GROMOS force-field [277]). The SETTLE algorithm [217] was used for keeping the bond length of water molecules at their equilibrium values, and the LINCS algorithm [216] was used to keep all remaining bonds constrained. The non-bonded interactions were calculated using a twin range method [192] with short and long-range cut-offs of 8 and 14 Å, respectively. A reaction field correction for truncated electrostatic interactions [204, 278] was applied, considering a dielectric constant of 54 [293]. The time step for integrating the equations of motion was 0.002 ps and the neighbor lists were updated every 5 steps.

The ATP-bound MalFGK$_2$E complex (hereafter designated **2ATP**) was first energy minimized to remove excessive strain. Initially, we performed 5000 steps of steepest descent minimization with harmonic restraints (with the force constant of 1000 kJ mol^{-1} nm^{-2}) applied to all heavy atoms, followed by another 5000 steps of the same algorithm, only restraining the protein heavy atoms, ending with 5000 steps with restraints applied to the Cα atoms only. After the minimization procedure, and in order to allow proper repacking of the lipids around the protein, we performed 500 ps of MD simulation with all protein, nucleotides and cofactors atoms harmonically restrained (the force constant used was 1000 kJ mol^{-1} nm^{-2}), at constant temperature and pressure. Afterwards, 250 ps of MD simulation were calculated, with position restraints (the force constant used was 750 kJ mol^{-1} nm^{-2}) applied to the protein non-hydrogen atoms only. Finally, only the Cα atoms were restrained (the force constant used was 500 kJ mol^{-1} nm^{-2}) for a period of 250 ps. The **2ATP** unrestrained simulations started after these 1-ns restrained simulations.

After 10 ns of unrestrained MD simulations for the **2ATP** state (see Figure 37), two new systems were built by transforming the original ATP nucleotides in ADP+IP (hereafter designated **2ADP.IP**) or in ADP (hereafter designated **2ADP**), using the “slow growth” method [294] to convert ATP into ADP or ADP+IP in a 5 ps time

interval. This process was performed using the same conditions as described for the long molecular dynamics simulations.

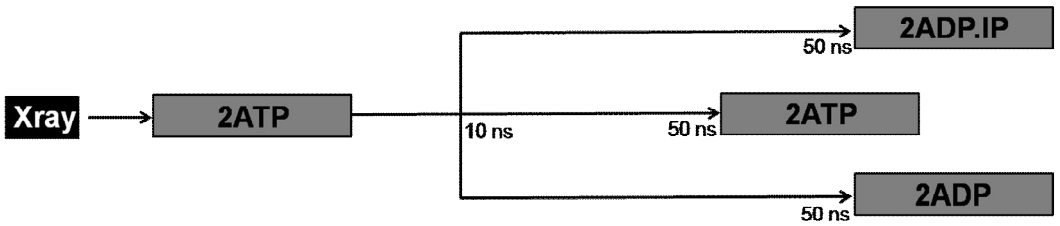


Figure 37- Outline of the MD simulations performed. Starting from the X-ray structure [62], the MalFGK$_2$E complex with MgATP bound in the binding sites was reconstructed and the E159Q mutation reverted in order to obtain the wild type complex. Then, the ATP-bound state (**2ATP**) was simulated for a period of 50 ns. After 10 ns of simulation of the **2ATP** state, two new states were prepared by replacing the ATP nucleotides by ADP (**2ADP**) or ADP and IP (**2ADP.IP**). Ten replicates, 50 ns each, were calculated for all the three states reported. This figure represents the procedure used for only one replicate.

In order to reduce the sampling problems in protein simulations, ten MD simulations, 50 ns each, were calculated for each state, in a total of thirty simulations, resulting in 1.5 μs of total simulation time. All replicates were initiated with different sets of random velocities.

5.3.4. Data analysis

All average structures were computed after Cα least-squares fitting to the X-ray structure, and by neglecting the first 40 ns of simulation.

The atomic positional deviations were calculated by comparing the different nucleotide bound states (**2ADP.I**P and **2ADP**) to the **2ATP** systems, within the

same replicate, and averaging these differences over all replicates (as described in [1]).

The secondary structure assignment was done using the DSSP program [262]. To determine the percentage of secondary structure loss relative to the X-ray structure, the regular secondary structure classes considered were the α-helix, the β-sheet, the 3_{10} helix and the β-bridge (DSSP classification). Only the residues that remain in the same secondary structure class as in the X-ray structure are counted.

5.4. Results

In order to investigate the conformational effect of ATP-hydrolysis and inorganic phosphate exit in an ABC importer system, we used as the starting point for our MD simulations, the ATP-bound outward facing conformation of the maltose importer, $MalFGK_2E$ [62]. This complex was simulated in the **2ATP**, **2ADP.IP** and **2ADP** states. The **2ADP.IP** state simulates the effect of two simultaneous ATP-hydrolysis, while the **2ADP** state mimetizes the simultaneous exit of both IP molecules from their respective binding sites.

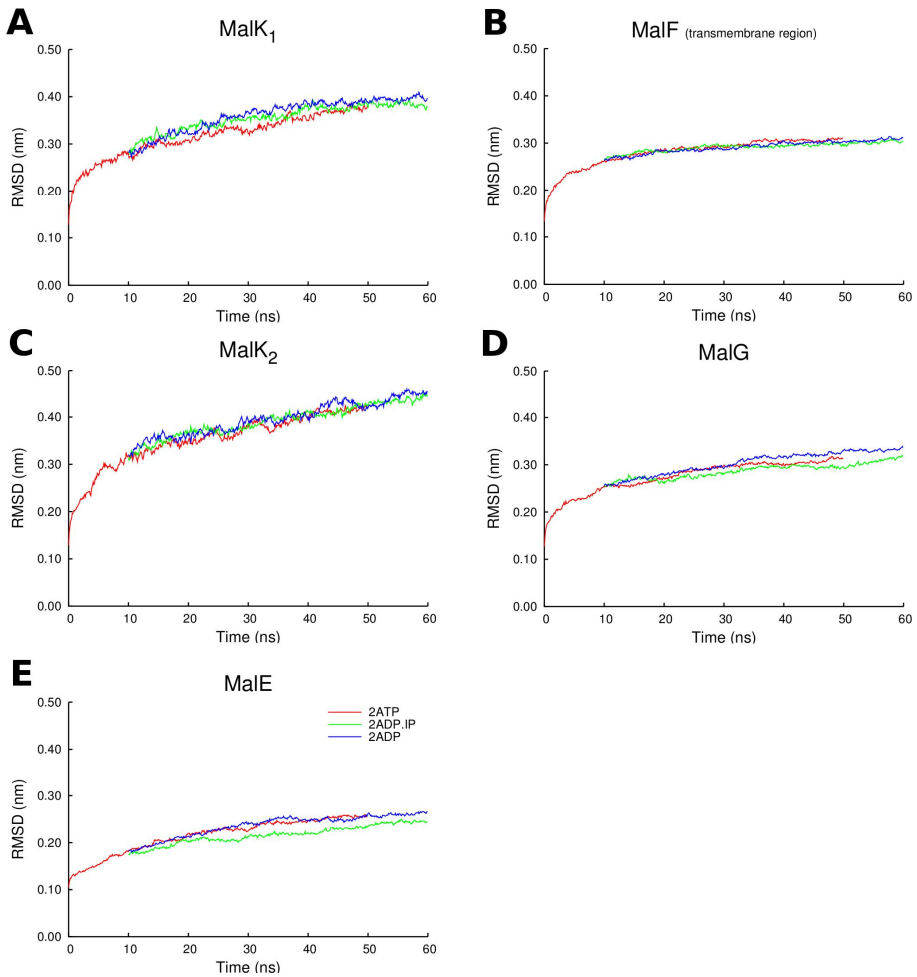


Figure 38- Conformational drift of the MalFGK$_2$E complex relative to the X-ray structure. Cα RMSD evolution along the simulation time for (**A**) MalK$_1$, (**B**) MalF transmembrane region. Due to a high amplitude rigid body rotation of the MalF periplasmic region (which will be debated in the following section), the RMSD was only determined for the MalF transmembrane region. (**C**) MalK$_2$, (**D**) MalG and (**E**) MalE. The Cα RMSD was calculated relative to the X-ray structure andwas determined after fitting each domain separately. Additionally, the RMSD values reported were averaged for all ten replicates for each state. Each point in the plot represents the average RMSD value for 100 ps.

The protein stability in our simulations was examined by visual inspection and by following the time evolution of several system properties, such as the root mean square deviation (RMSD) from the X-ray structure (see Figure 38) and the percentage of retained native secondary structure (SS) (Figure 57 of the Supporting Information section 10.3).

As can be seen in Figure 38, the RMSD profiles for all MalFGK$_2$E components increased continuously in the **2ATP** state, reaching the global values of about 0.31 nm for the transmembrane domains (MalF and MalG) and 0.25 for MalE. The MalK domains exhibit more deviations from the X-ray structure as the Cα RMSD increases continually until it reaches values of 0.37 nm and 0.41 nm for MalK$_1$ and MalK$_2$, respectively. In general, the RMSD values reported, although large, are in the same order of magnitude as the ones reported in previous MD studies for other ABC transporters [162, 164, 168]. The evolution of the native SS content (see Figure 57 of the Supporting Information section 10.3) show that the overall secondary structure remains intact, with a secondary structure loss lower than 12% at the end the 50 ns simulation of the **2ATP** state**.** The RMSD values, together with the time evolution of native SS content, are indicative of structural stability for the simulated complexes.

5.4.1. Rigid body motion of the MalF Periplasmic region

In the first 20 ns of the **2ATP** simulations (in all 10 replicates) we observed a rigid body rotation of large amplitude towards the membrane, in the periplasmic region of MalF, namely in the MalF-P_2 region (also designated by some authors "MalF-P_2 loop" [62]), which resulted in an high RMSD for this region (see Figure 58 of the Supporting Information section 10.4). However, the overall tri-dimensional fold of this transporter is maintained intact during the simulations (see Figure 39 for a representative example).

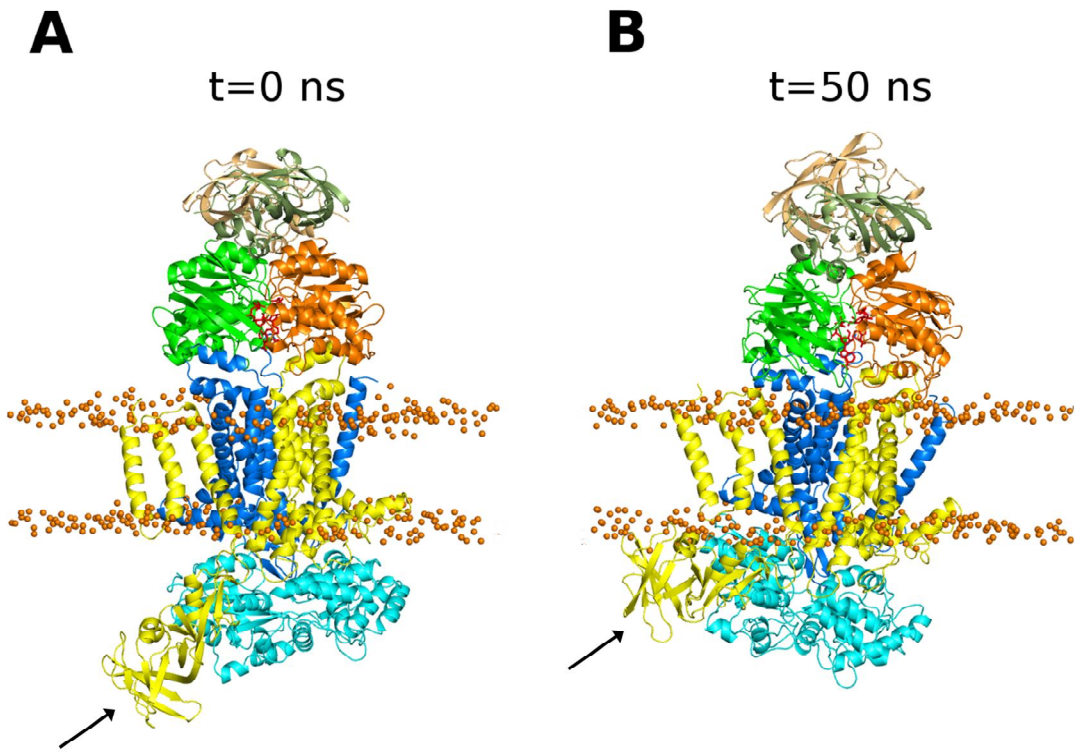


Figure 39- Detail view of the MalF-P_2 region (residues 91-271) movement in the **2ATP** replicate 3 simulation. $MalK_1$, $MalK_2$, MalF, MalG and MalE are colored is green, orange, yellow, blue and cyan, respectively. In these images, the regulatory domains are colored with lighter colors. ATP is shown as red sticks, whereas the phosphate atoms from the lipid head groups are represented as orange spheres. **A**- Snapshot at the beginning of the simulation. **B**- Structure obtained after 50 ns of

simulation. The zone experiencing this rigid body movement is marked with a black arrow in both figures.

After this large amplitude movement, the MalF-P_2 region side chains establish new interactions with the membrane lipid head-groups. Furthermore, together with the MalF-P_2 region upward movement, the MalE N-terminal region is pulled toward MalG. The MalG-MalE approximation observed in our ATP simulations, is in agreement with experimental evidences reporting that, both MalF and MalG, are involved in MalE binding during the transport cycle [295, 296], a fact that was not observed in the crystallographic structure [62]. Biochemical and genetic studies suggested that the MalE N-terminal lobe interacts mainly with MalG, whereas the MalE C-terminal region is close to MalF during the transport cycle [295, 296]. Our simulations seem to capture the functional details of this ABC transporter.

5.4.2. MalK dimer interface

In a previous MD simulation work, Wen and Tajkhorshid [96], using the isolated NBDs subunits of the maltose transporter (MalK) in solution, were able to observe the atomic details of the MalK interface opening upon ATP hydrolysis, in a timescale ranging from 30 to 50 ns. In order to see if the MalK dimer dissociation was observable in our complete transporter simulations, the distance between the P-loop of one MalK monomer and the ABC signature motif of the other MalK monomer was determined for all states (see Figure 59 in the Supporting Information section 10.5). From this measurement, it is evident that in the **2ATP** and in the **2ADP.IP** states, no MalK dimer interface separation was detected in all ten replicates, within the 50 ns simulated, evidencing that there was no dissociation of the MalK dimer. However, in the **2ADP** state, in one of the ten replicates (replicate 8), the dissociation of the MalK interface, in binding site 2, was observed after 37 ns of MD simulation (see Figure 40 and Figure 59 in the Supporting Information

section 10.5). In this replicate, at the beginning of the simulation, both binding sites are closed and the two ADP molecules are bound at the interface between monomers. At the end of the simulation, binding site 2 is completely separated and the nucleotide is only bound to the P-loop residues. This observation seem to agree with the experimental evidences showing that, in the MalK dimer, ADP, unlike ATP, cannot promote ATPase dimerization nor the stabilization of the closed dimeric form [53].

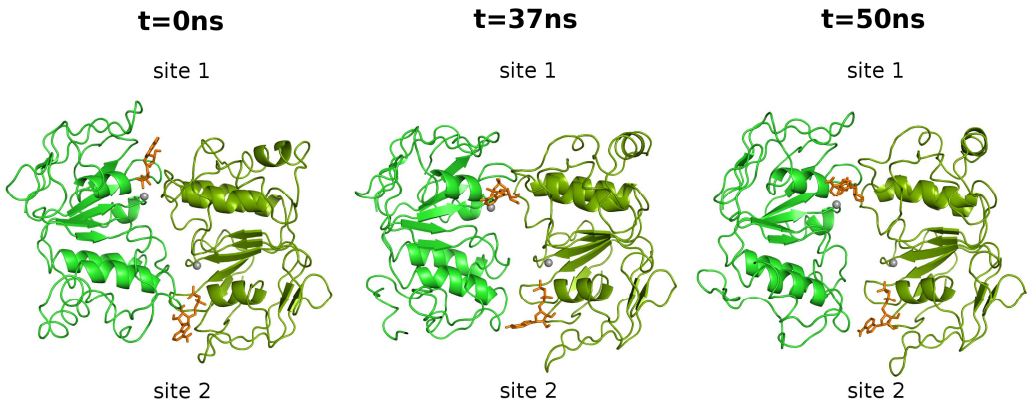


Figure 40- Snapshots of the MalK dimer opening in the **2ADP** state for replicate 8. The left side image was obtained at 0 ns, whereas the middle represents the beginning of the dimer dissociation (after 37 ns). The right side image corresponds to the final structure (after 50 ns). The MalK monomers are colored in green, ADP is represented with orange stick and Mg with gray spheres. For clarity purposes, the regulatory domains are not represented in these three images.

In the ADP-bound state, with the objective of analyzing with more detail the MalK interface dissociation and the rearrangements involved in this process, several distances were measured for replicate 8. To quantify dimer separation, we determined the distance between the P-loop and the ABC signature motif forming the binding sites (Figure 41A). Additionally, the distances between the ADP nucleotide and the P-loop (Figure 41B), and the ABC signature motif (Figure 41C), were also determined.

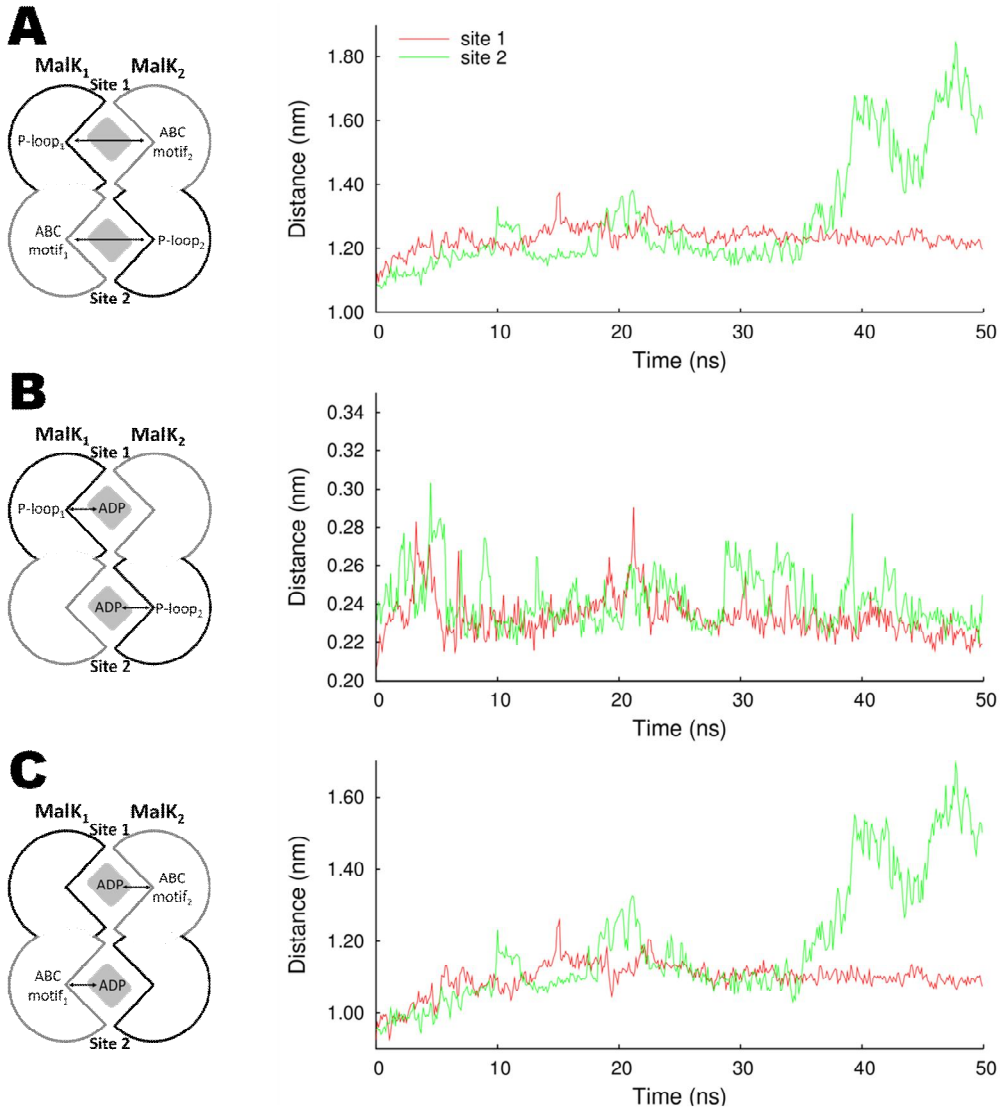


Figure 41- Opening of the MalK interface in the **2ADP** state (replicate 8). The red and green lines corresponds to distances measured in binding site 1 (formed by the $MalK_1$ P-loop and the $MalK_2$ ABC motif) and 2 (formed by the $MalK_2$ P-loop and the $MalK_1$ ABC motif), respectively. Each point represents the average value over 100 ps. **A**) Distance between the P-loop motif of one MalK monomer and the ABC signature motif of the other. **B**) Distance between the ADP α,β-phosphate and the P-loop residues. **C**) Distance between the ADP α,β-phosphate and the ABC signature motif residues.

As can be seen in Figure 41A, during the MalK interface dissociation, the distance between the binding site residues increases ~0.44 nm in binding site 2, whereas in binding site 1 this distance does not change considerably. Moreover, the interface opening is accompanied by the dissociation of the ADP phosphate atoms from the ABC signature motif (Figure 41C), although they did not dissociate from their corresponding P-loop residues (Figure 41B), preventing the nucleotide diffusing away from the binding site. Similar observations for NBD dimer dissociation were reported for the simulations of isolated MalK dimer [96] and for the ADP.IP-bound Sav1866 exporter [2]. Surprisingly, and although ADP is present in the two binding sites, only site 2 presents a clear opening of the MalK interface. This asymmetric movement, observed in previous theoretical works (not only in the MalK dimer [96], but also for several other ABC members [2, 56, 97]), is frequently attributed to the stochastic nature of the opening process. This cannot be clarified here, since we have only one event in the ten replicates simulated.

Despite this considerable motion in the NBDs, the effect in the transmembrane domains is very limited, and no major conformational changes were observed in MalF and MalG, within the simulated timescale (see Figure 60 of the Supporting Information section 10.5).

5.4.3. Structural differences during the ATP-cycle in the MalFGK$_2$E complex

It is nowadays accepted that the ATP-cycle (nucleotide binding/hydrolysis and release of its products) induces conformational changes in the ATPase domains, which are then transmitted to the transmembrane region [59, 286]. In order to identify the hydrolysis-dependent conformational changes and the residues involved in inter-domain communication, the Cα positional deviation (relative to the **2ATP** state) was calculated for both post-hydrolysis states (Figure 42). These deviations were determined as a function of the residue number, for the last 10 ns

of simulation, using a methodology previously described for the identification of the hydrolysis-dependent conformational rearrangements in the isolated MJ0796 NBD dimer [1].

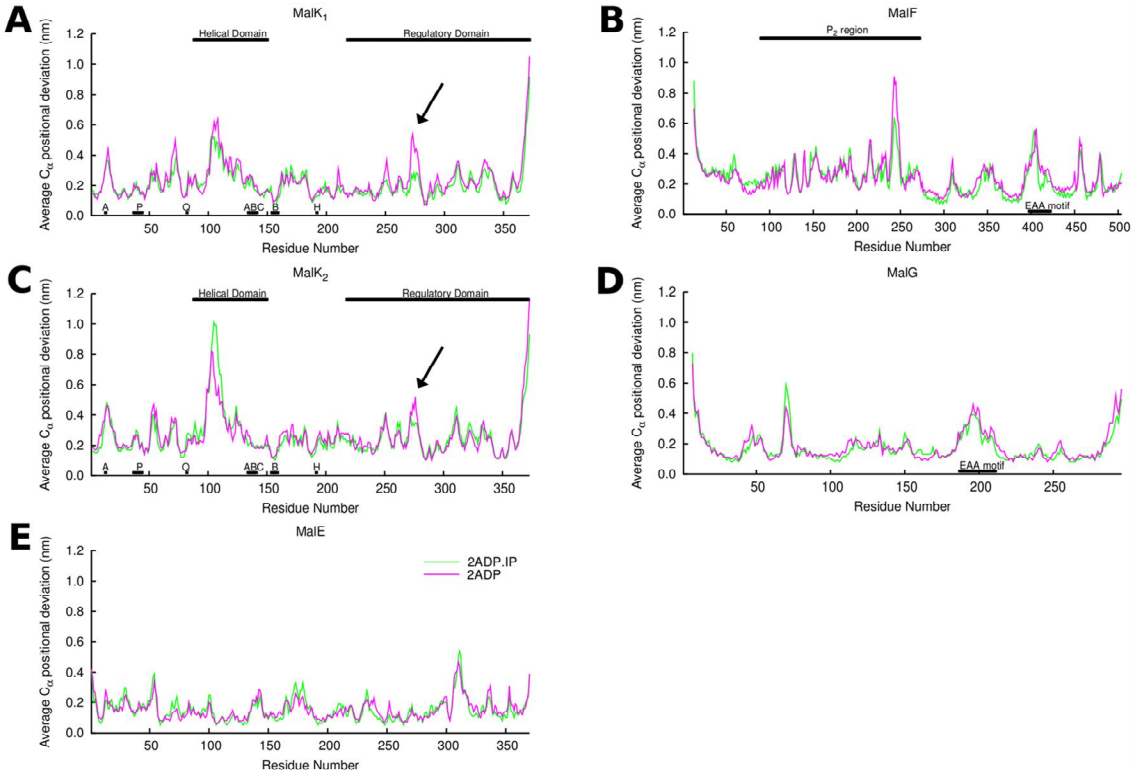


Figure 42- Average Cα Positional Deviation (over the last 10 ns of the simulations) with respect to the average **2ATP** conformation. The deviation values displayed were obtained by comparing the average structure of each replicate with the average **2ATP** structure for the same replicate, in order to eliminate the structural differences arising from the natural variation between replicates (for more details about this methodology see [1]). The individual differences were averaged afterwards over all ten replicates. The relevant structural family motifs are marked in the plots: P-loop (P), Q-loop (Q), ABC signature motif (ABC), Walker-B motif (B), the H-motif (H) and the "EAA" motif (where the Coupling Helix is located [60]). In figure **A** and **C**, the black arrows identify the segment ranging from residue 271 to 276, which undergoes a significant rearrangement after IP exit. Average Cα Positional Deviation for **A**-$MalK_1$, **B**- MalF, **C**- $MalK_2$, **D**-MalG and **E**-MalE.

Upon analysis of Figure 42A and Figure 42C, we can observe that, although the regions showing the largest conformational changes are generally the same in the two MalK monomers, the amplitude of these changes is significantly different between monomers. The clearest example of this MalK asymmetric behavior is the HD region, where the conformational changes are more pronounced in the $MalK_2$ monomer. This uneven behavior has already been reported in previous MD works, not only in the MalK dimer [94, 96], but also in other ABC members [1, 2, 87, 93, 95, 98, 167, 170]. At this stage, we are still unable to conclude whether this asymmetry arises from the X-ray structure itself (due, for example, to different crystallographic contacts) or if it is an intrinsic characteristic necessary for the functioning of this protein. In the present case, the MalK asymmetry can also arise from the effect exerted by the TMDs in the ATPase domains, which, contrary to the NBDs, are two different membrane proteins. For $MalK_1$ and $MalK_2$, the segments 14-15, 99-112 and the C-terminal (residues 366-371) can be identified as the regions presenting relevant differences between the two post-hydrolysis and the pre-hydrolysis state. In general, the regions presenting higher Cα positional deviation values are the same for **2ADP.IP** and **2ADP**. However, one exception was detected in the segment 271-276 (marked by a black arrow in Figure 42A and Figure 42C), which undergoes significant rearrangement after IP exit only. For the transmembrane domains, MalF and MalG (Figure 42B and Figure 42D), the major conformational deviations, relative to the **2ATP** state, are concentrated in very narrow and specific regions. In MalF, the zones presenting relevant rearrangements during the ATP-cycle are the segments 240-249 and 402-407, and the N-terminal. In MalG, the residues with higher positional deviations are the 70-72 and 190-197 segments, the N-terminal and the C-terminal regions. Finally, and based on the low average positional deviation values observed in Figure 42E, MalE seems to be very little affected by nucleotide hydrolysis and IP exit (with the exception of the residues ranging from 310-311), at least in the 50 ns timescale.

In order to identify the spatial organization of the residues undergoing the most relevant displacements during the ATP-cycle, the position of these residues were mapped in the average structure, both for the **2ADP.IP** and the **2ADP** states (Figure 43).

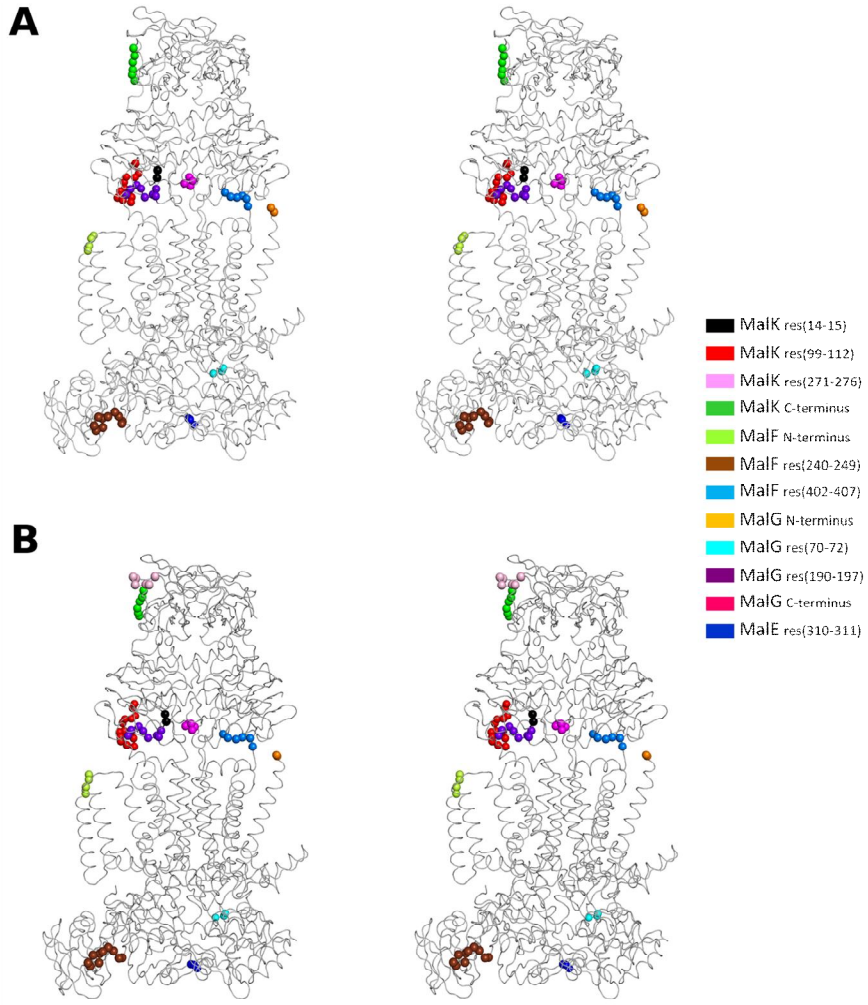


Figure 43- Stereo image of the global average structure for the **A**) **2ADP.IP** and **B**) **2ADP** states. The global average structures were determined from the last 10 ns of simulation, over all ten replicates, after fitting to the X-ray structure. The regions presenting the most relevant positional deviations, relative to the average **2ATP** state, are highlighted in different colors (see the legend). For clarity, only

the $MalK_1$ segments are colored, but similar positional deviations are also observed in the same $MalK_2$ segments.

From Figure 42A, Figure 42C and Figure 43, one of the major conformational changes observed in the NBD region is located in the 14-15 segment (also named A-loop). The A-loop is located in the external region of the MalK monomer, connecting two β-sheets (S1 and S2). This region was named A-loop due to presence of an essential conserved aromatic residue, which is known to interact with the nucleotide adenine ring during the ATP cycle [77]. Our previous work [1], performed in an the isolated ATPase dimer from *Methanococcus jannaschii*, has also identified the A-loop segment as one of the regions showing major conformational changes during the ATP-cycle [1]. The other MalK region evidencing high Cα deviation is the HD region (residues 99-112). This region is located in an external and flexible region of the MalK monomer, and is in direct contact with the transmembrane domains (MalF and MalG), more specifically with the coupling helices. Several previous works (both theoretical [1, 2, 87, 98] and experimental [35, 36, 51, 86]) have also identified the HD has one of the regions more affected by the ATP-hydrolytic cycle. Additionally, the HD has long been thought to be involved in inter-domain communication and energy transmission from the NBDs to the TMDs [297], mainly throughout an outward rotation movement. In our simulations, we could determine a small amplitude rotation of the HD upon hydrolysis (see Figure 61 of the Supporting Information section 10.6), although the observed rotation is significantly smaller than the ones previously described in simulation works performed on the isolated NBD dimers [1, 87, 97, 98] or inferred from X-ray analysis [35, 36, 51-53, 86]. This difference may be related with the timescale of our simulations and/or to the effect exerted by the TMDs (especially the “EAA motif”) in the HD region. It is also possible that the absence of the TMDs in previous simulation works allowed the observation of non-physiological, higher amplitude, movements in the HD.

In MalF (see Figure 42B and Figure 43), two segments are mostly affected by nucleotide hydrolysis and IP exit. The first segment is formed by residues 240-249 and it is located in the MalF-P_2 region, directly contacting MalE. The second segment comprises residues 402-407 and is located in the "EAA" motif, more specifically in the coupling helix region. The "EAA" motif (EAA-X(3)-G) [67, 68] is the direct contact point between MalK and MalF/G, and it is formed by two short cytoplasmic helices oriented roughly parallel to the membrane plane [62]. One of these helices, named Coupling Helix [60], docks directly into a cleft in the helical sub-domain region of MalK [62], providing the bulk of the interdomain contacts. Despite the limited sequence similarities between the TMDs region in the ABC family [298], other ABC members (such as the Sav1866 exporter [58] and the HI1470/71 importer [63]) also present coupling helices located in similar regions, and directly interacting with the ATPase domains. Moreover, over the last years, it has been suggested that these helices may have an essential role in the mechanism of transmission of conformational changes between domains [58].

In MalG (Figure 42D and Figure 43), the regions showing major conformational changes during the ATP-cycle are the residues 70-72, 190-192 and the C-terminus. The first set of residues is located in the periplasmic side of the membrane, in direct contact with MalE, while the second set is located in the "EAA" motif, in the coupling helix region (similarly to MalF). During the ATP-cycle, the two coupling helices (one in MalF and the other in MalG), although different in sequence, are both similarly affected by the corresponding MalK HD rearrangements. Lastly, the MalG C-terminal region is also highly influenced by the ATP-cycle, mostly due to its location, since it is inserted into the MalK dimer interface, close to the Q-loop motifs (close enough to form some hydrogen bonds with the Q-loop residues) [62]. It was also suggested that, although the MalG tail is not essential for the NBDs dimer formation [52], the interactions formed between these two domains contribute for the Q-loop ordering, and may be import for the intermediate states of the transport cycle [62]. The Q-loop motif, and especially, the conserved glutamine residue located in this

region, has long been suggested to play a key role in hydrolysis, by coordinating the magnesium ion located in the binding site, and by orienting the nucleophilic water needed for ATP hydrolysis [51].

5.4.4. Maltose Position during ATP hydrolysis

In the X-ray structure, the maltose binding site is located at the base of the transmembrane cavity, approximately halfway across the membrane and it is formed exclusively by MalF residues [62]. Maltose is bound by hydrogen bonds and by ring stacking interactions to several MalF aromatic residues [62]. During our simulations, the maltose position and the allocrite binding site rearrangements were monitored in order to understand if hydrolysis and the presence of IP directly influences the binding site conformation, and consequently the allocrite affinity for this site. This was done for all replicates of the **2ATP**, **2ADP.IP** and the **2ADP** states. For the majority of the replicates, maltose does not move away from the binding site (see Figure 62 in the Supporting Information section 10.7), but there are two exceptions, one in the **2ATP** (replicate 2) and another in the **2ADP.IP** (replicate 1) simulations. In these two cases, the interactions with the MalF binding site residues (see Figure 63 in the Supporting Information section 10.7), mainly the hydrogen bonds, are severely reduced (data not shown), which enables maltose to move laterally and exit from the binding site, interacting with some MalG residues (as for example, glutamine 129 and glutamate 229). Nevertheless, in all replicates, no significant change was observed during the 50 ns of simulation in the overall shape of the transmembrane cavity and in the maltose position between the three distinct states (see Figure 64 of the Supporting Information section 10.7).

5.5. Concluding Remarks

Although ABC transporters have been widely studied during the last thirty years, the understanding of this family and its transport mechanism is still very incomplete. The main objective of this work was to identify and characterize the major conformational rearrangements occurring during an ATP-cycle in a model import system ($MalFGK_2E$), using theoretic techniques. In order to achieve our goal, extensive MD simulations (50 ns long, with 10 replicates) on the full-length $MalFGK_2E$ complex were performed for three distinct states of the ATP-cycle: a pre-hydrolysis (with two ATP molecules bound to the binding sites), a post-hydrolysis state (with ADP and IP bound) and a post-IP exit state (with only ADP bound). These simulations allow us to gain new relevant insights into inter-domain communication mechanism and into the conformational rearrangements induced by the ATP-hydrolytic cycle. First of all, we show that the structural changes induced by the nucleotide hydrolysis and IP exit are restricted to specific residues or segments, both in the ATPase domains (MalK) and in the transmembrane regions (MalF and MalG). The segments presenting major rearrangements are the A-loop and the HD region, for MalK, and the “EAA motif” region and the coupling helix, for both MalF and MalG. Additionally, the MalG C-terminal region is also severely affected by hydrolysis, mainly due to its location (inserted between the two MalK monomers). The MalE domain, surprisingly, is generally not affected by hydrolysis, at least within the simulated timescales. Given the conserved minimal ABC architecture and the similarities of the NBDs across the entire ABC transporter super-family, it seems reasonable to expect a common mechanism for coupling hydrolysis to allocrite translocation, both for importers and exporters. The MD simulations performed here for the $MalFGK_2E$ importer, and reported previously for the Sav1866 exporter [2], allowed the identification of the HD region and the “coupling helices” as the headstones of inter-domain communication in the ABC family.

In the three studied states (**2ATP**, **2ADP.IP** and **2ADP**), we were able to observe a high amplitude rigid body rotation of the MalF-P_2 region towards the membrane, together with the approximation of the MalE N-terminal region to the transmembrane domain MalG. This MalE-MalG approximation correlates well with the experimental evidences showing that the MalE N-terminal mainly interacts with MalG during a transport cycle.

In one replicate in the **2ADP** state, we were able to observe the MalK dimer interface dissociation, in one of the two binding sites. In this case, and similarly to what was described in other simulation works [2, 96, 97], the interface opening was accompanied by the dissociation of ADP from the ABC signature motif, but not from the P-loop motif. Surprisingly, the opening of one of the nucleotide binding sites did not significantly alter the MalF and MalG conformations, at least in our simulated timescale.

We were also able to observe that ATP hydrolysis does not affect the allocrite binding site, and in the majority of the replicates (28 of 30 replicates), maltose does not abandoned this site during the whole simulation.

Over the last years, molecular dynamics studies (performed both in complete importers and exporters [2] as well as in the isolated NBD dimers [1, 87, 96, 97, 172]) have intensively contributed with new insights into the communication mechanism in ABC transporters. A general mechanism for coupling hydrolysis and energy transduction to allocrite translocation (independently of the transporter directionality) in ABC transporters is now starting to be unraveled and its “key players” are the HD region and the TMD coupling helices.

5.6. Acknowledgements

This work was supported by a FCT (Fundação para a Ciência e Tecnologia) fellowship (SFRH/BD/21433/2005).

6. General Discussion

Over the last decades, the continuing growth of computing power has made possible to tackle new and more complex biological problems using theoretical methodologies. Computer simulations of biomolecules are now frequently used to obtain new atomic insights into biology and substantially improve our understanding of protein structure, function and dynamics. Thanks to the current computational capabilities together with new searching algorithms and improved force-fields, classic all atom molecular dynamics simulations can realistically describe the dynamic behavior of a protein at time scales that were unthinkable only a few decades ago.

As stated in the introduction, due to the high sequence and structural similarities of the ATPase domains throughout the whole ABC transporter family, it has been hypothesized a general mechanism for ATP binding, hydrolysis and communication with the TMDs common to all ABC members, independently of the directionality of the transport and the nature of the allocrites. In this thesis, we describe several MD simulation works performed in model ABC transporters with the main objective of gaining new insights into the functioning of this family. Three distinct transporters were studied: the isolated MJ0796 NBD dimer, the Sav1866 exporter and the $MalFGK_2E$ importer. The principal subject that crosses all the main chapters of this thesis refers to the structural consequences of nucleotide hydrolysis and to the identification of the interdomain communication mechanism. The first work reported in this thesis (chapter 3 and [1]) uses the isolated NBD dimer of a lipid transporter (MJ0796) [188] found in the hyperthermophilic Archaeon *Methanococcus jannaschii* [80, 188, 189]. The second work reported (chapter 4 and [2]) uses the complete Sav1866 lipid exporter from *Staphylococcus aureus* [58, 59], which is a homologue of clinically relevant multidrug exporters [58]. The third work (chapter 5 and [3]) simulates one of the best functionally characterized ABC importers, the maltose importer ($MalFGK_2E$) from *E. coli* [62, 142].

Isolated MJ0796 NBD dimer

ABC transporters couple ATP-hydrolysis to allocrites transport throughout membranes. In order to identify the immediate consequences of ATP hydrolysis (< 30 ns) and its conformational rearrangements in the ATPase domains, we used MD simulation techniques to study the isolated MJ0796 NBD dimer in several potential intermediate states of the ATP-hydrolytic cycle. Overall, five distinct states, each containing different nucleotides (ATP, ADP and the inorganic phosphate species) in the binding sites, were simulated. From these five states, we were able to have a clearer picture of the conformational changes occurring in an isolated NBD dimer and identify the segments involved the NBD-TMD communication mechanisms, mainly the A-loop and the helical sub-domain (HD) region. The first segment, the A-loop, is located in a peripheral region of the dimer, positioned just above the adenine ring of the nucleotide, and it contains a conserved aromatic residue involved in π–π stacking interactions with the adenine base of the nucleotide. The second segment, the HD, is located in an external and flexible region of the dimer and it is in direct contact with the TMDs. Additionally, from our simulations we were also able to suggest a new possible exit path from the binding sites for the IP species, mainly due to an outward rotation of the loop comprised of residues 38-43. As this segment moves away from its initial position, the IPs becomes totally exposed to the solvent, increasing its probability of diffusing away from the binding sites prior to the NBD dimer interface dissociation.

In terms of future perspectives, we think it would be interesting to extend the MJ0796 simulations in order to try to observe the NBD dimer interface opening in the post-hydrolysis states. Additionally, the nucleotide release, as well as the IP diffusion from the binding sites are also events that require longer simulation times. Moreover, other intermediate states could be studied in order to clarify different questions, such as the dynamics of nucleotide driven NBD closure. Several theoretical studies have already addressed the NBD closure [56, 94, 171], but we

can go beyond these works mainly by ensuring proper statistics for the MD simulations, by performing long nanosecond timescale MD simulations and by using a considerable number of replicates.

Sav1866 exporter

As already stated, it is currently proposed that nucleotide hydrolysis induces conformational changes in the NBDs, which are, by a still unknown molecular mechanism, transmitted afterwards to the TMDs. Several theoretical works performed in the isolated NBD dimers [1, 87, 96, 97, 172], have already gave us insights into the NBD-TMD signal transmission events. However, in complete transporters these events can be significantly different from the ones observed previously in the isolated NBD dimers (mainly due to the effect exerted by the TMDs). Therefore, the main objective of the second work (chapter 4 and [2]) was to identify the ATP-hydrolysis induced conformational changes in a complete ABC exporter and compare the results with the ones previously observed for the isolated NBD dimers [1, 87, 96, 97, 172]. For that purpose, we performed extended nanosecond timescale MD simulations of the complete Sav1866 exporter inserted into a lipid bilayer in two distinct states of the ATP-cycle: a pre-hydrolysis (ATP-bound) and a post-hydrolysis state (ADP and inorganic phosphate-bound). From these simulations, we were able to characterise the initial conformational changes that are a consequence of ATP hydrolysis, in both the TMDs and the NBDs. We were also able to show that the effects of hydrolysis are not restricted to the binding sites or to the NBDs, but extend deeply to the external regions of the TMDs, showing that the mechanism of energy transduction in ABC transporters is a rapid and complex process. The regions mostly affected by hydrolysis are the extracellular loops 1 and 3, coupling helices 1 and 2 in the TMDs and the HD region in the NBDs. Additionally, in the post-hydrolysis state simulations, it was also possible to observe, for the first time, the NBD dimer interface dissociation within the context of a complete transporter. These simulations gave us new and relevant insights into the binding site opening mechanism and they showed that the binding

site dissociation is accompanied by the ABC signature motif separation from ADP and IP, and it is a consequence of ATP-hydrolysis, rather than the dissociation of the hydrolysis products. Moreover, in both studied states, we were able to observe the opening of the cytoplasmic gate entrance located in the ICL region, which exposes the allocrite transmembrane pathway to the cytoplasm. In order to study the permeation properties of the transmembrane pathway towards a known allocrite, doxorubicin, we determined the minimum adiabatic energy profiles along the transmembrane pore. Curiously, the energy profiles for both states are overall downhill from the cytoplasmic to the extracellular side, which can explain, at a molecular level, the unidirectionality in allocrite translocation, and the power stroke in transport. The cytoplasmic gate entrance was identified as the only allocrite controlling zone, and significative differences in the energy profiles between the pre- and the post-hydrolysis states were detected. The energy barrier located in the cytoplasmic gate region shows a large reduction upon hydrolysis, allowing efficient allocrite permeation. It is important to refer that the results reported in this work may have important implications in our understanding of the mechanism of allocrite transport and its associated power stroke.

Having in mind that none of our MD simulations included the allocrite molecule, in the future, it could be interesting to perform new longer MD simulations with the allocrite molecule inside the transmembrane pathway, in order to see if any difference upon hydrolysis would be detected. Additionally, a detailed study of the interactions formed between the allocrites and the transmembrane pathway residues is still required. Moreover, it would also be interesting to simulate the mechanism of substrate translocation through the transmembrane pathway, determine the rate limiting zones and see the effect of hydrolysis in other ABC exporters (such P-gp [65]). We think it should be exciting to compare those results with the work presented here.

MalFGK$_2$E importer

As described previously, the structures for exporters [58, 59, 64, 65] and importers [57, 60-63, 66, 142, 250] are significantly different, mainly in the transmembrane region, which can suggest a different interdomain communication mechanism. Having this hypothesis in mind and in order to identify and characterize the largest conformational changes induced by the ATP-cycle in a complete ABC importer, we performed several extensive MD simulations of the complete MalFGK$_2$E importer inserted into a lipid bilayer in three distinct states: a pre-hydrolysis (ATP-bound), a post-hydrolysis (ADP and inorganic phosphate-bound) and a post-IP release state (ADP-bound) (chapter 5 and [3]). From the simulations, we showed that the major structural changes are located in the A-loop and in the HD region, for MalK, and in the coupling helices, for MalF and MalG. Moreover, the MalG C-terminal region is also severely affected by hydrolysis, mainly due to its location between the two MalK monomers. In the MalFGK$_2$E case, and similarly to what was observed for the Sav1866 exporter [2], the energy from hydrolysis is converted into conformational changes in the HD region, which are transmitted from the NBDs to the TMDs via the coupling helices. At this point it seems that a unified inter-domain communication mechanism and coupling of ATP hydrolysis to transport is starting to appear and that the NBD-TMD signal transmission events are common to all ABC transporters. Moreover, in the MalFGK$_2$E simulations we were able to detect, in the post-IP release state, a MalK dimer interface dissociation similar to the ones observed previously in isolated NBD dimers [96, 97] and in the Sav1866 exporter [2]. This observation seems to suggest that the NBD dimer dissociation mechanism, like the inter-domain communication mechanism, is common to all ABC transporters. Curiously, the NBD dimer interface opening did not significantly affect the MalF and MalG conformation and no considerably change in the transmembrane pathway was identified, most probably due to the short timescale of the simulations.

As future work in the ABC importers field, and in particular in the $MalFGK_2E$, it could be interesting to study the permeation of the allocrites across the transmembrane pore in different nucleotide bound states, similarly to what was done for the Sav1866 exporter [2]. Additionally, we could also use non-standard MD techniques (such as Umbrella Sampling [190]) to characterize the free energy profiles along the channel in different states of the ATP-cycle. Lastly, in order to generalize our observations to all ABC importers, we should at least perform similar MD studies in other model ABC importer, such as the Vitamin B_{12} transporter [57].

7. General Conclusion

As the reader can attest, in this thesis we present exhaustively the molecular modeling studies performed in the ABC transporter family done by us during the last four years. Our major goal, during this PhD, was to clarify how is the ATP energy harnessed from the nucleotide binding sites and transmitted from the catalytic to the transmembrane domains. From all MD simulations (overall ~80 distinct simulations for both the isolated NBD dimers and for the complete transporters were performed), we were able to identify the conformational response to hydrolysis and the regions responsible for NBD-TMD communication transmition. Moreover, our simulations also suggest that the NBD dimer dissociation mechanism is common to all ABC transporters, and the molecular details of the opening are similar for both importers and exporters. Lastly, we have also contributed with relevant insights to the understanding of how the allocrite transport is “powered” in ABC exporters. In conclusion, we believe that the work presented in this thesis was able to provide a detailed molecular description of the ABC transporters mechanism of action and helped to clarify some of the oldest unanswered questions in this family. Additionally, we also clearly showed that theoretical methods are a valuable tool to study biological phenomena that are difficult to be addressed experimentally.

8. Supporting Information for chapter 3

8.1. ATP, ADP and IP Parametrization

8.1.1. Protonation state of Adenosine Triphosphate (ATP), Adenosine Diphosphate (ADP) and Inorganic Phosphate (IP)

In the physiological pH range, the overall charge of ATP may be -3 or -4 and the charge of ADP -2 or -3 (see the protonation equilibrium Figure 44) in the absence of a metal counter ion. Therefore it is important to know which protonation states are present at the catalytic pocket of the protein. Both molecules have several possible protonation positions and the protonation/ionization equilibrium in aqueous solution is well described in the literature [299]. Additionally, the presence of the counter ion bound to the protein can influence the proton binding equilibrium. The formation of Mg^{2+}ATP is reported to reduce the pK_a values by 2.0-2.8 units [300-302].

Based in experimental data available (the X-ray dimer in ATP bound form [51]) and in the use of theoretical methodologies [239, 255] the thermodynamic equilibrium of proton binding was studied at pH=7.0. The theoretical methodologies used combine continuum electrostatics, calculated with the package MEAD (version1.1.8) [234, 240, 256], and Metropolis Monte Carlo simulations in the proton binding space, using the program PETIT [239].

The results show that both chemical species (ADP and ATP) in the complexed form are deprotonated in the oxygens of the P_G and in the N_1 atom of the adenosine and are protonated in the O_{2*} and O_{3*} of the sugar. The overall charge is -3, -2, -2 and -1 for the ATP, ADP, the complex formed between Mg^{2+}ATP and Mg^{2+}ADP, respectively.

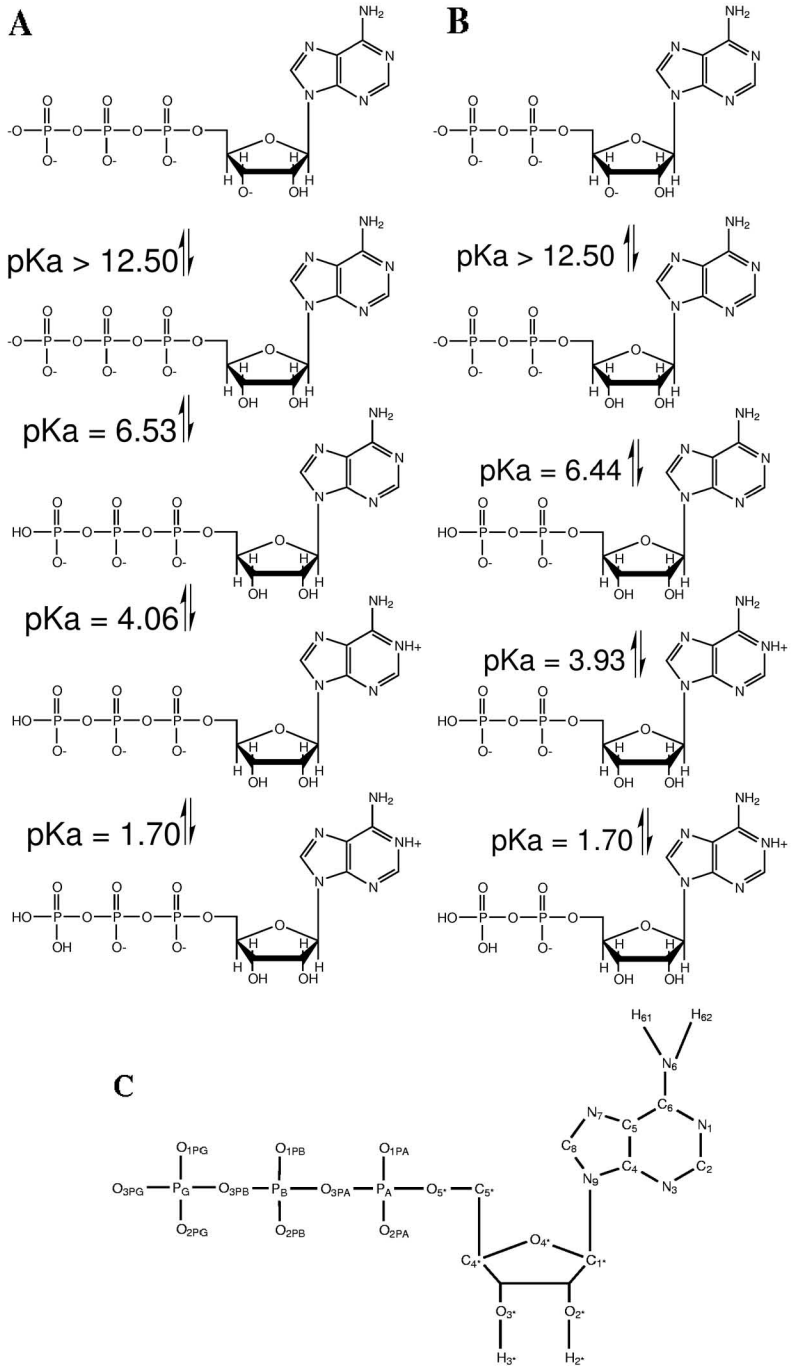


Figure 44- Ionization and protonation equilibrium of ATP (**A**) and ADP (**B**) in aqueous solution. C- shows the atomic names of ATP taken from the GROMOS96 force field [196, 199].

The protonation equilibrium for the IP ion (PO_4^{3-}) in aqueous solution is also well known [303]. At physiological pH the anion may be single or double protonated, with charge -2 or -1, respectively. We chose the single protonated state, due to the presence of a positively charged ion (Mg^{2+}) in the proximity of nucleotide. In this case, the introduction of another positive charge would be energetically disfavoured by the presence of magnesium.

8.1.2. Parameterization of the ATP, ADP and IP

Given that the GROMOS 43A1 force field only contains parameters for the protonated versions of ATP and ADP, we had to derive partial charge distributions for their relevant deprotonated versions. For calculating partial charges for ATP and ADP we follow the approach adopted by Meagher et al. [304] and used the Hwang et al. extended conformations [305] in quantum chemical calculations. The partial charges for the phosphate group atoms (plus a CH_3 group introduced at the end of the phosphate chain in order to mimic the connection to the nucleoside) of ATP and ADP were calculated with the program GAUSSIAN98 [306] and RESP fitting [282]. Electrostatic potentials calculated with the 6-31+G(d) basis using RHF and geometry optimization were used for fitting of partial charges using the RESP procedure. Contrary to the usual procedure, polarization functions were used to account for a highly charge molecule (as done by Meagher et al. [304]). The nucleotide charges (up to C_5*) were the same as the ones used in the protonated ATP in the 43A1 GROMOS96 force field. The overall charge is -4 and -3 for ATP and ADP, respectively.

The atomic charges for the inorganic phosphate (IP) were also calculated with the same procedure described above. The overall charge for the inorganic phosphate (HPO_4^{-2}) is -2.

The values of the partial charges used for the atoms are listed in **Table 5**, for ATP, ADP and IP:

Table 5- Partial charges for the ATP, ADP and IP species. Both ATP and ADP are deprotonated at the oxygens of P_G and P_B, respectively. The calculation of the charges for both ATP and ADP considered the atoms from the O_{5*} up to oxygens of the phosphate groups.

Atom	ATP	ADP
O_{5*}	-0.436	-0.485
P_A	1.752	1.739
O_{1PA}	-1.038	-1.030
O_{2PA}	-1.038	-1.030
O_{3PA}	-0.769	-0.692
P_B	1.658	1.716
O_{1PB}	-1.006	-1.072
O_{2PB}	-1.006	-1.072
O_{3PB}	-0.527	-1.072
P_G	1.732	-
O_{1PG}	-1.108	-
O_{2PG}	-1.108	-
O_{3PG}	-1.108	-

Atom	IP
O_1	-1.127
P_2	1.927
O_3	-0.949
H_4	0.404
O_5	-1.127
O_6	-1.127

8.2. Comparison between ATP/ATP and E171Q-ATP/ATP simulations: structural Equivalence between the mutant and the wild type dimer

It is important to note that the X-ray structure chosen as starting point for this study is a mutant and may not represent the exact details of the natural physiological dimer. So, to begin with, we needed to be certain of the structural similarity between the mutant and the wild type dimers. Several molecular dynamics simulations of 5 ns each were performed for both the mutant and the wild type systems. In all replicates we considered the first 2 ns as an equilibration period. Analysis of all individual simulations showed that the root mean square deviation (RMSD) from the initial structure is mostly stabilized after these initial 2 ns (Figure 45).

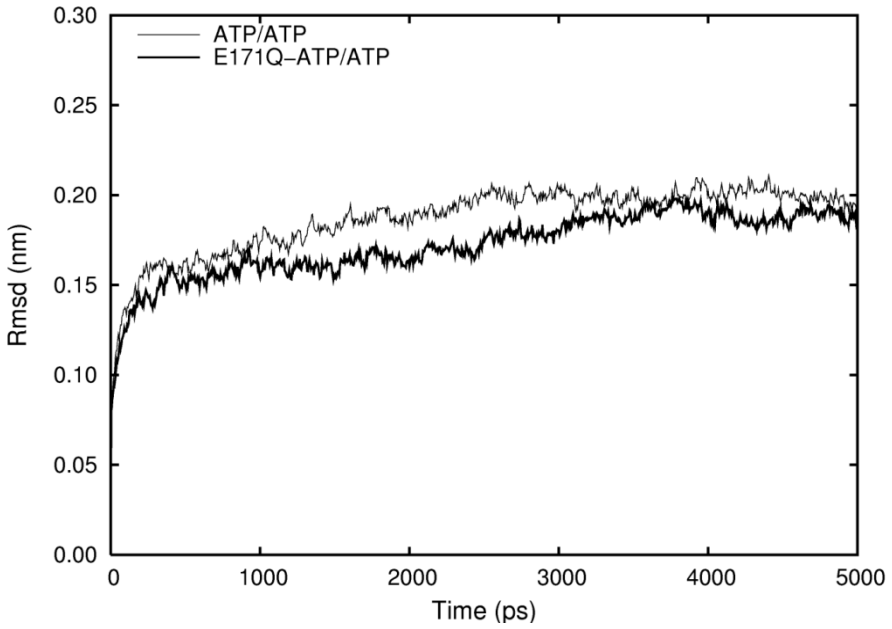


Figure 45- Time evolution of the C-α RMSD from the X-ray dimer structure in the simulation. RMSD averages (obtained after fitting to the X-ray structure) for all replicates are used here.

After the stabilization period, the conformations resulting from the simulations show a global RMSD from the crystal structure of about 0.18 nm for the wild type and 0.17 nm for the mutant.

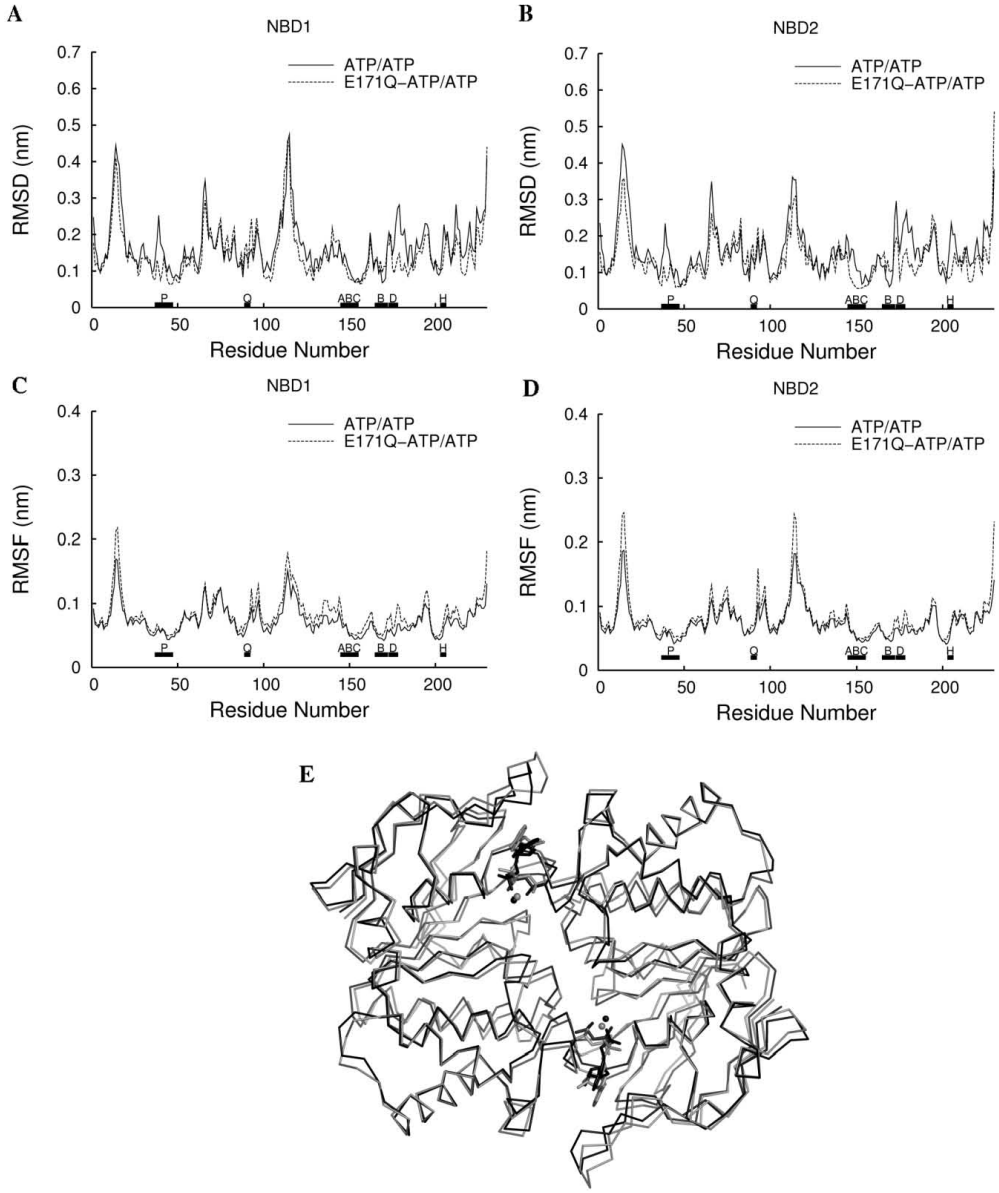


Figure 46- Figure A and **B**- Average C-α RMSD of **ATP/ATP** (black line) and **E171Q-ATP/ATP** (dotted line) with respect to the crystal structure for the NBD_1 (**Figure 46A**) and NBD_2 (**B**). The RMSD

values displayed were calculated by comparing the average structure for each replicate (obtained after fitting to the crystallographic structure) to the X-ray structure and only afterwards these values were averaged over all replicates. Also, the RMSD was calculated by superimposing simultaneously both NBD_1 and NBD_2 to the crystal structure. The labels shown in the graphic correspond to the position (in the polypeptide chain) of the ABC structural family motifs: P-loop (P), Q-loop (Q), ABC signature motif (ABC), Walker-B motif (B), D-loop (D) and H-motif (H). **Figure C** and **D**- Averaged Cα RMSF of **ATP/ATP** (black line) and the **E171Q-ATP/ATP** (dotted line) over the last 3 ns of all the replicates for both NBD_1 (**Figure 46C**) and NBD_2 (**D**). The labels in this graphic have the same meaning as in Figure 24. **Figure E**– Comparison between the average structure of **ATP/ATP** dimer (black line) and **E171Q-ATP/ATP** dimer (grey line) structures, along the last 3 ns of all replicas. All average structures were obtained after fitting to the crystal structure. In both cases ATP is represented with sticks while the cofactor is shown as spheres. The cofactor is Mg^{2+} for the **ATP/ATP** dimer and Na^{+} for the **E171Q-ATP/ATP** dimer. Figures generated with the program PyMOL [252].

In order to characterize more precisely the structural differences between the wild type **(ATP/ATP)** and the mutant **(E171Q-ATP/ATP),** we compare the Cα RMSD and the Cα fluctuations (RMSF) for each residue as well as the average three-dimensional structure in order to identify the location of the most variable regions between them (Figure 46). From the results in Figure 46, we can conclude that both **ATP/ATP** and **E171Q-ATP/ATP** have a similar behaviour during the MD simulations and that the residues displaying higher RMSD and RMSF are the same in the two systems. These observations led us to assume that the mutation of the glutamate 171 to a glutamine does not involve relevant conformational changes in the overall protein structure, even though it inactivates the complex. These results led us to consider the two systems structurally similar and validate the use of the X-ray structure to study the wild type dimer. For these reasons we proceeded the study of the wild type dimer (**ATP/ATP**) only.

8.3. Time evolution of the Cα RMSD for all replicates.

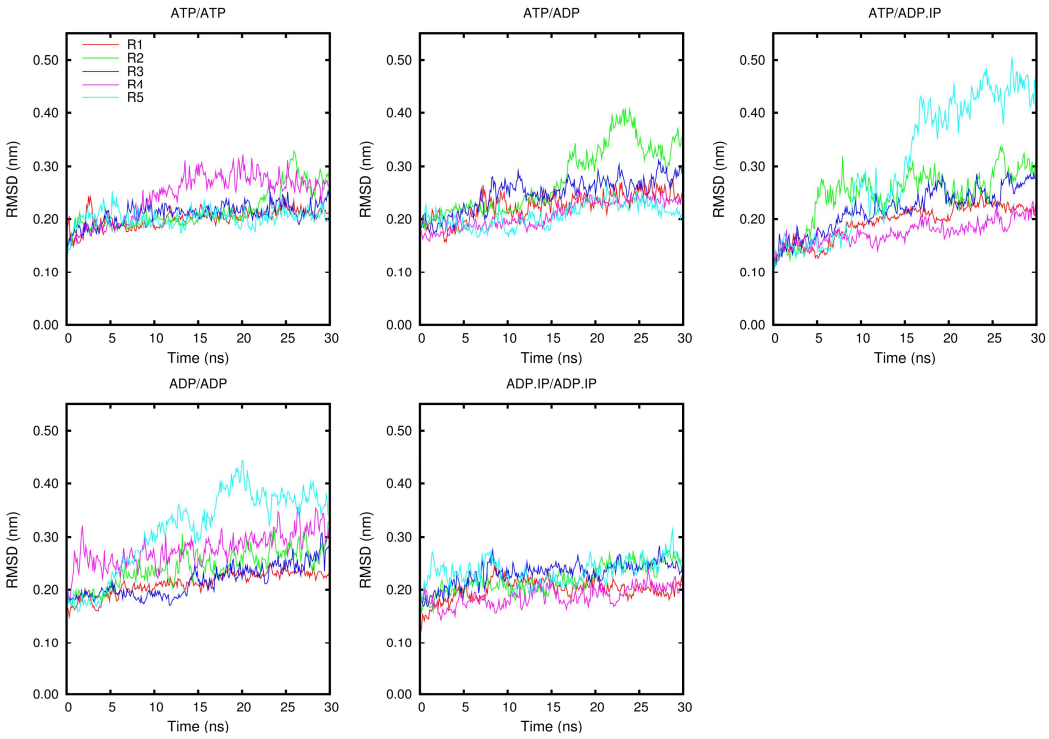


Figure 47- Time evolution of the Cα RMSD relative to the X-ray structure (fitting to this X-ray structure) for all 5 replicates simulated for all the states reported in this study (**ATP/ATP**, **ATP/ADP**, **ATP/ADP.IP**, **ADP/ADP** and **ADP.IP/ADP.IP**). Moreover, each point represents the average RMSD value over 100 ps.

From Figure 47, we can observe that after 30 ns of simulation all the systems are more or less equilibrated for the majority of the replicates.

8.4. Average Cα positional Deviation relative to the ATP/ATP conformations for all 25 possible pairs of simulations.

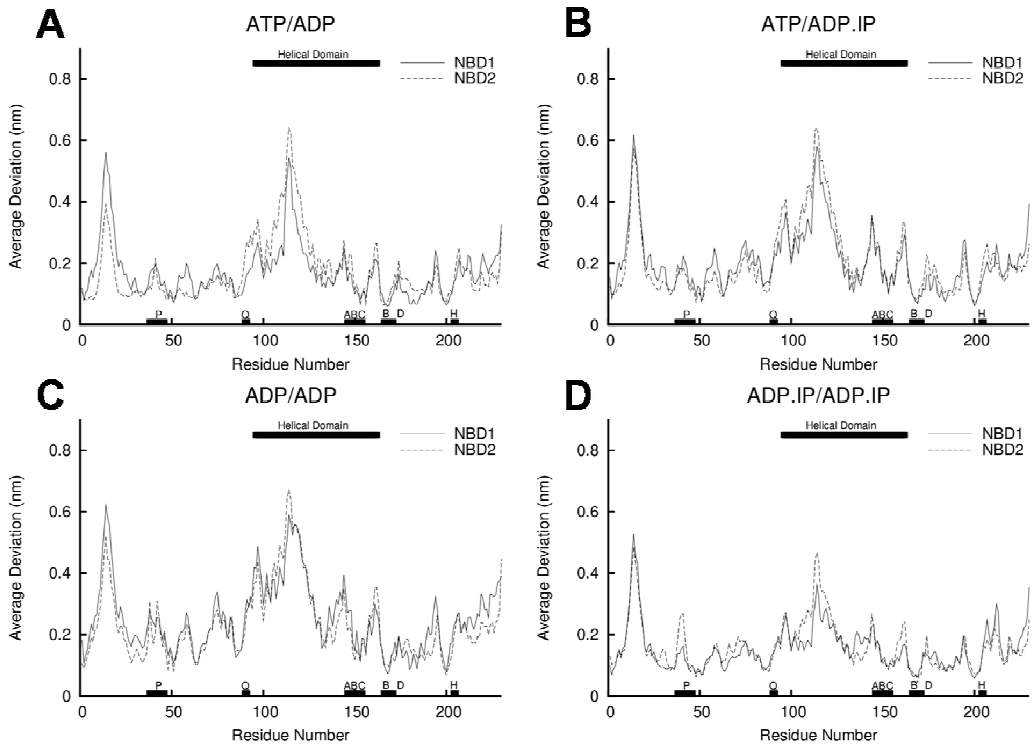


Figure 48- Average Cα Positional Deviation (over the last 10 ns of the simulations) relative to the **ATP/ATP** conformation for: **A**) **ATP/ADP**, **B**) **ATP/ADP.IP**, **C) ADP/ADP** and **D**) **ADP.IP/ADP.IP**. The deviation values displayed for the different nucleotide bound states were computed by comparing the whole ensemble of conformations for one state with the whole **ATP/ATP** conformational ensemble. The results were then averaged over all combinations. The graphs show the position (in the polypeptide chain) of the ABC structural family motifs and some relevant sub-domain regions: P-loop (P), Q-loop (Q), ABC signature motif (ABC), Walker-B motif (B), D-loop (D), H-motif (H) and Helical Domain (residue 96-162).

9. Supporting Information for chapter 4

9.1. Protonation state of protonable residues

The protonation state of each individual protonable group in the protein and in the nucleotides species at a given value of pH (in this case pH=7.0) has to be specified prior to the MD simulations. We determined these protonation states using methodologies for studying the thermodynamics of proton binding described elsewhere [239, 255]. These methodologies use a combination of Poisson-Boltzmann calculations, performed with the package MEAD (version 2.2.0) [234, 240, 256], and Metropolis Monte Carlo simulations, using the program PETIT (version1.3) [239]. Since Sav1866 is a membrane protein with an asymmetrical transmembrane pore, a custom made model for the transmembrane pore and corresponding calculation was built (see Figure 49).

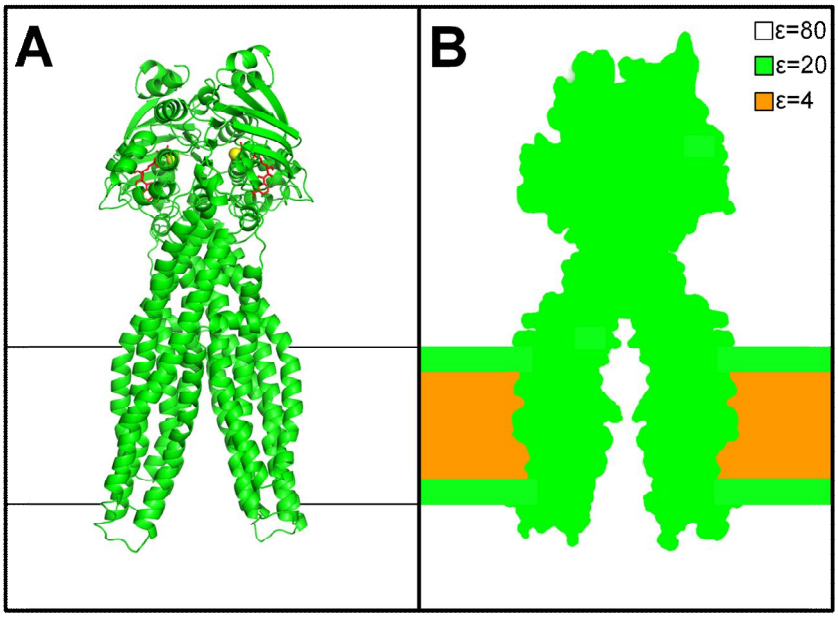


Figure 49- A schematic view of the model used to calculate the protonation states with the program MEAD (version 2.2.0) [234, 240, 256] and PETIT (version 1.3) [239] for the

Sav1866 protein inserted in a lipid bilayer. **A**- Overview of the crystal structure of the ABC transporter Sav1866. The protein is rendered as a cartoon, the ATP molecules are represented by red sticks and the magnesium ions as yellow spheres. **B**- Overview of the complete model for the Sav1866 transporter built by the package MEAD. The membrane is represented by three slabs of low dielectric media. The first and third slabs mimetize the hydrophilic lipid head groups of each leaflet of the bilayer (ε=20), and the middle slab (coloured in orange) represents the lipid hydrophobic tails (ε=4). In this model, the slabs representing the hydrophilic parts of the lipids are 7.5 Å wide, whereas the slab representing the lipid tails is 26 Å wide.

In the Sav1866 model, the membrane was introduced in the x-y plane as a low dielectric media formed by tree different regions (see Figure 49B). The first region mimetizes the hydrophilic lipid head group, the second region characterizes the middle of the bilayer representing the hydrophobic tails of the lipids and the third slab represents again the hydrophilic lipid head group of the opposite leaflet. Different values of dielectric constant (ε) were assigned based in the atomic z coordinates: for the first and the third slab of the bilayer the dielectric constant used was 20, while for the middle of the bilayer the ε was 4.

Moreover, a custom made pore was made through the membrane (see Figure 49B). In this case, all points located inside the transmembrane pore are treated with the dielectric constant of the solvent (ε=80).

According to our calculations, all glutamic and aspartic acids are deprotonated, while lysines and arginines are protonated, with the exception of Lys26 that has a total charge of zero. The N-terminal should be in the neutral state (NH_2) while the C-terminus was considered negatively charged (COO^-). All free cysteines are protonated. The protonation state of the histidines is described in Table 6.

Table 6- Protonation state of Sav1866 histidines. HisA refers to the residue with a zero total charge and protonated at the Nδ position, while HisB corresponds to a neutral residue protonated at Nε. The HisH represents the protonated histidine at both Nδ and Nε with the overall charge of +1.

His Number	Protonation State	
	Chain 1	Chain 2
48	HisB	HisB
57	HisB	HisB
58	HisB	HisB
103	HisB	HisB
204	HisB	HisB
240	HisA	HisA
344	HisH	HisH
403	HisH	HisH
457	HisB	HisB
534	HisH	HisH
541	HisH	HisH
552	HisA	HisA
559	HisB	HisB
571	HisH	HisH

9.2. DMPC bilayer construction

9.2.1. 128 DMPC bilayer

An initial model for a hydrated 128 DMPC bilayer (64 lipids in each leaflet and 4214 water molecules), obtained from Tieleman's website at http://moose.bio.ucalgary.ca, was used as a starting point for the construction of our DMPC membrane. The water/lipid content was about 33 water molecules per lipid.

Since the initial membrane model had been equilibrated under different simulation conditions, we submitted the 128 DMPC molecules to an extra 30 ns of MD simulation. The simulation was performed using the GROMACS 3.3.1 package [213, 214, 257, 307]. The lipid parameters used were based on the 53A6 GROMOS96 force field [196, 197, 199, 200] except for the atomic partial charges, which had been previously derived by Chiu and co-workers [275]. The membrane equilibration was performed at constant temperature conditions (310K). This temperature is above the phase transition temperature reported in literature for DMPC (Tm=296-297K) in order to ensure that the bilayer is in the liquid crystalline state [276]. The setup for the molecular dynamics simulations is similar to the one reported in the main article in the materials and methods section.

The evolution for some lipid properties (area and volume per lipid) along the simulation time can be followed in Figure 50.

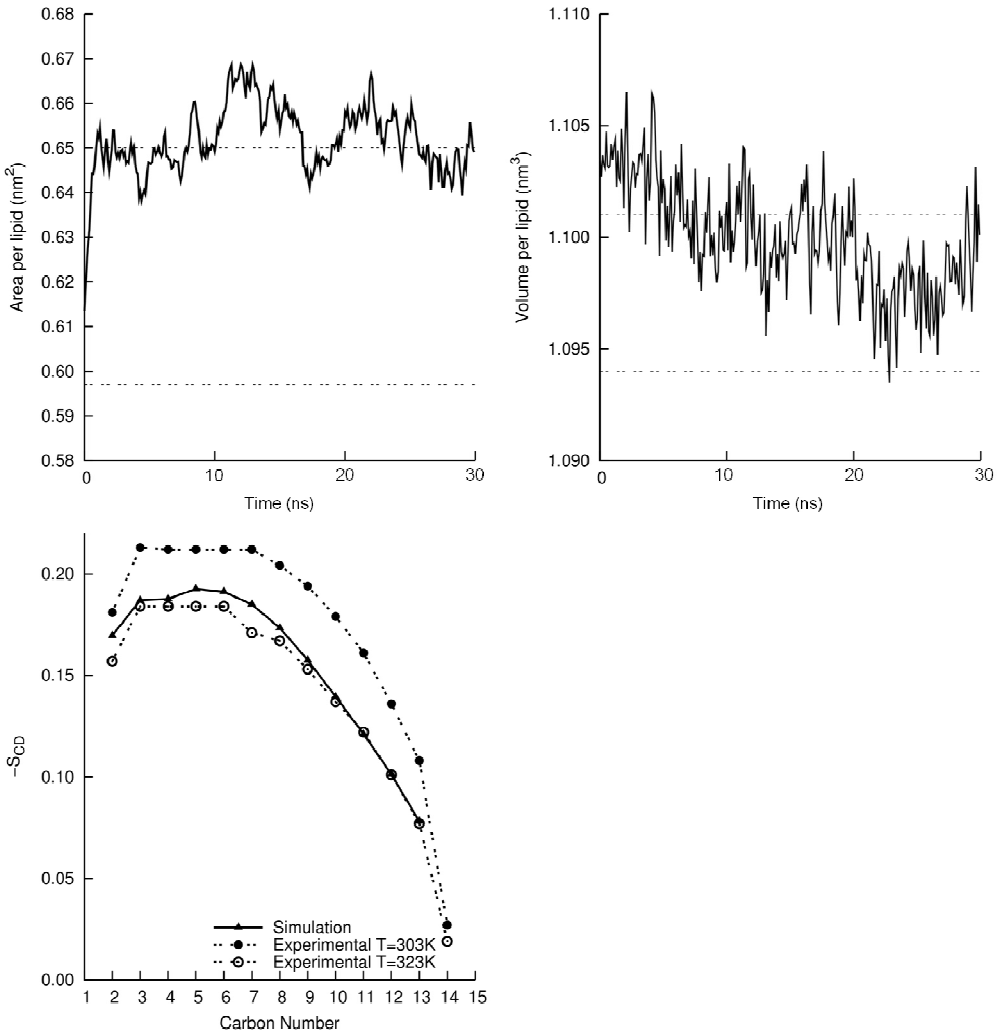


Figure 50- **A**- Area per lipid for the 128 DMPC as a function of the simulation time (solid black line). The horizontal dashed lines represent the upper and lower limit ranges for the experimental values [308]. **B**- Time evolution of the volume per lipid (solid black line). Again, the horizontal dashed lines represent the upper and lower limit ranges for the experimental values [308]. In both plots, each point represents the average value over 100 ps. **C**- Deuterium order parameters. Black and white circles are the experimental results [309] averaged over both sn-1 and sn-2 chains for two different temperatures (T=303K and T=323K) while the black triangles represent the data obtained from the present simulation.

As can be seen, in the end of the 30 ns of simulation, the area (Figure 50A) and volume per lipid (Figure 50B), despite experimenting some fluctuations, remain more or less within the experimental frame. The average values for the area and volume per lipid over the last 5 ns of simulation are 0.648 nm^2 and 1.098 nm^3, respectively. Based on previous works with DMPC lipidic bilayers [310], we considered the first 25 ns as an equilibration period and this data was discarded. Moreover, in the liquid crystalline phase (the biologically relevant phase for our study), the lipid tails are fluid and disordered. Additionally, the organization of the hydrocarbon tails can be calculated from the simulation and compared to experimental results obtained using NMR techniques (Figure 50C).

9.2.2. 512 DMPC bilayer

The 512DMPC bilayer was generated by the replication of the already described above pre-equilibrated 128 DMPC bilayer in both the x and y directions. The resulting system was then energy minimized and re-equilibrated again for over 20 ns. The system was subjected to 5000 steps of steepest descent energy minimization in order to remove excessive strain, and then submitted to 100 ps of restrained MD simulation: the first 20 ps with restraints in all heavy atoms, including water (force constant 1000 kJ mol^{-1} nm^{-2}), followed by 30 ps with restraints in all lipid heavy atoms (force constant 1000 kJ mol^{-1} nm^{-2}) and the last 50 ps with restraints in all the lipid heavy atoms, but in this case with a lower value for the force restraints (5 kJ mol^{-1} nm^{-2}). The unrestrained MD simulations started from this point onward and lasted 20 ns. All the simulation details were exactly like the ones described in section 4 in the materials and methods. At the end of the re-equilibration period, similarly with the 128 DMPC bilayer, the 512DMPC bilayer presents area and volume per lipid values within the experimental frame. The average value for the area and volume per lipid over the last 5 ns of simulation is 0.658 nm^2 and 1.099 nm^3, respectively.

9.3. Doxorubicin Parameterization

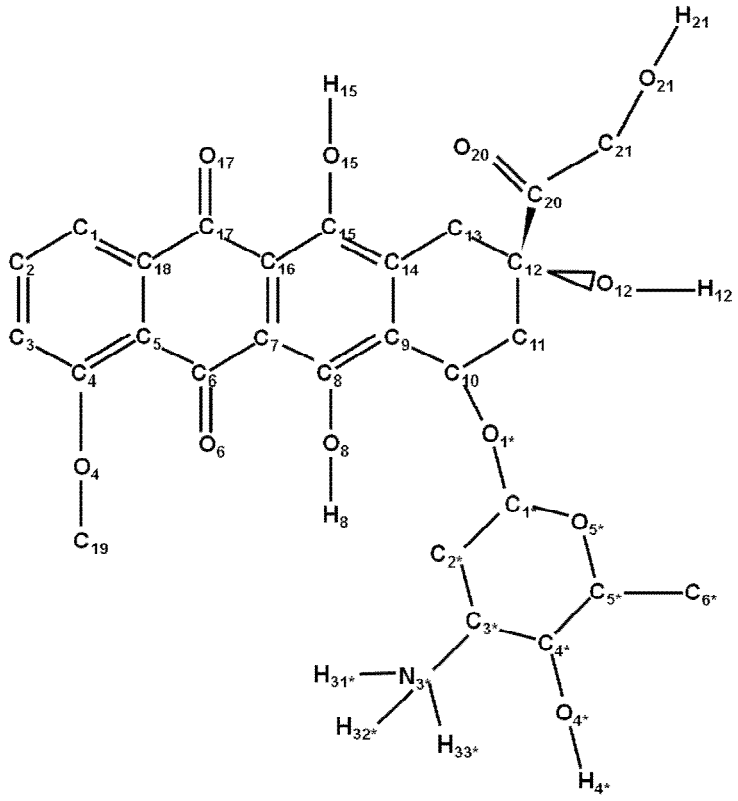


Figure 51- Atomic names for the doxorubicin molecule. In this parameterization only the polar hydrogens were considered explicitly.

Given the large conformational space available to doxorubicin and the fact that the sugar and the group starting on atom C_{20} can freely rotate in relation to the rest of the molecule (see Figure 51), a multiconfigurational RESP fitting procedure (using four conformations) was used [282] for the determination of the atomic partial charges of this molecule. The first conformation corresponds to the X-ray structure [311] whereas two other result from 180 degrees rotations applied to the sugar and to the group starting in atom C_{20}. Finally, a fourth conformation containing the two

described rotations simultaneously was also considered. All four conformations were optimized at the Hartree-Fock level with the 6-31G(d) basis set, using the Gaussian 03 program [281]. Electrostatic potentials were obtained for the four conformations and used to derive united atom charges in a two stage RESP fitting procedure. In the first stage all charges were freely fitted, except the ones from the C_6 and C_{17} atoms, which were constrained to a previous fitting where aromatic protons were considered. This was done because their charges, in an unconstrained fitting, were found to be unphysically high. In a second stage, we constrained the charges of the protons belonging to the NH_3 group to be the same, and refitted charges again, while keeping carbon charges fixed in the results of the first fitting. The partial charges calculated for the doxorubicin atoms are listed in Table 7:

Table 7- Partial charges for Doxorubicin. This molecule is considered positively charged in the daunosamine group at pH=7.0.

Atom	Charge	Atom	Charge
C1	0.223	C17	0.695
C2	0.090	O17	-0.406
C3	0.156	C18	-0.734
C4	-0.263	C19	0.323
O4	-0.223	C20	0.367
C5	0.266	O20	-0.481
C6	0.709	C21	0.346
O6	-0.619	O21	-0.681
C7	-0.432	H21	0.419
C8	0.455	O1*	-0.334

O8	-0.596	C1*	0.298
H8	0.473	C2*	-0.016
C9	-0.341	C3*	0.462
C10	0.375	N3*	-0.743
C11	0.048	H31*	0.402
C12	0.259	H32*	0.402
O12	-0.654	H33*	0.402
H12	0.380	C4*	0.282
C13	0.143	O4*	-0.766
C14	0.051	H4*	0.482
C15	0.045	C5*	0.159
O15	-0.450	O5*	-0.385
H15	0.349	C6*	0.083
C16	-0.019		

9.4. Protein stability during the MD

Ten replicates were simulated for the two systems reported in this study (**2ATP** and **2ADP.IP**). The protein behaviour was carefully examined for all replicates during the 60 ns of the simulation, whether by visual inspection or by the monitorization of several system properties, such as the number of H-bonds between the two monomers, the root mean square deviation (RMSD) from the X-ray structure and the solvent accessible surface (SAS). In several replicates, both in the **2ATP** or the **2ADP.IP** systems, significant deformations in the proteins tertiary structure were observed, especially in the TMDs which lead to the structure bending towards the membrane. This can be easy understood upon observation of Figure 52, where the last conformation of the simulation for each replicate is represented.

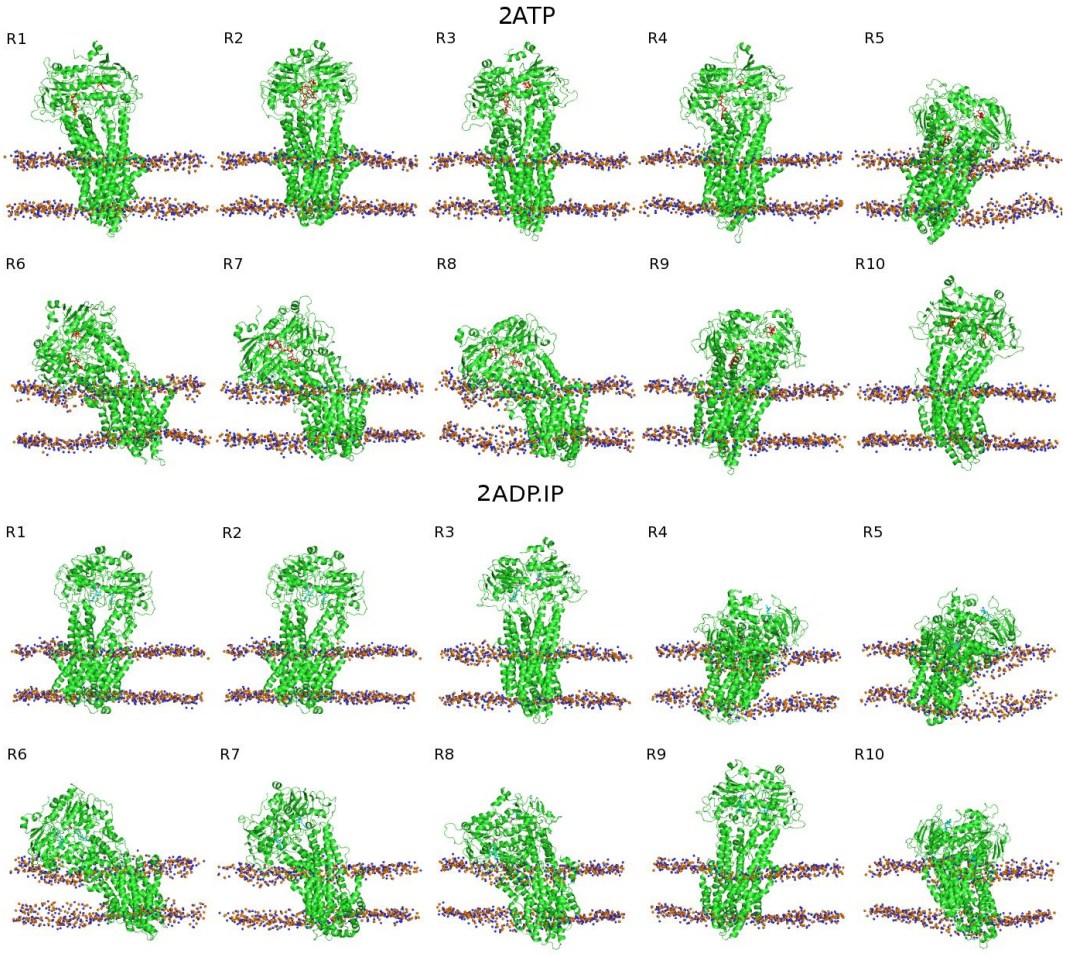


Figure 52- Snapshots of the Sav1866 transporter after 60 ns of simulation for the ten replicates performed for the **2ATP** and for the **2ADP.IP** systems**.** The protein is rendered as a carton while the nitrogen and the phosphate atoms from the lipid head groups are represented as blue and orange spheres.

For this reason, for the structural comparison between the ATP- and the ADP.IP-bound states, we decided to analyse only the replicates that were less deformed in both states (replicate 1, 2 and 3). This choice allowed us to investigate the hydrolysis-induced conformational changes without having to handle with the high

amplitude bending motions. The remaining seven replicates (replicates 4 to 10) were not considered in the following structural analysis**.**

9.5. Relation between canonical Q-loop glutamine-Mg interaction and the helical sub-domain rotation

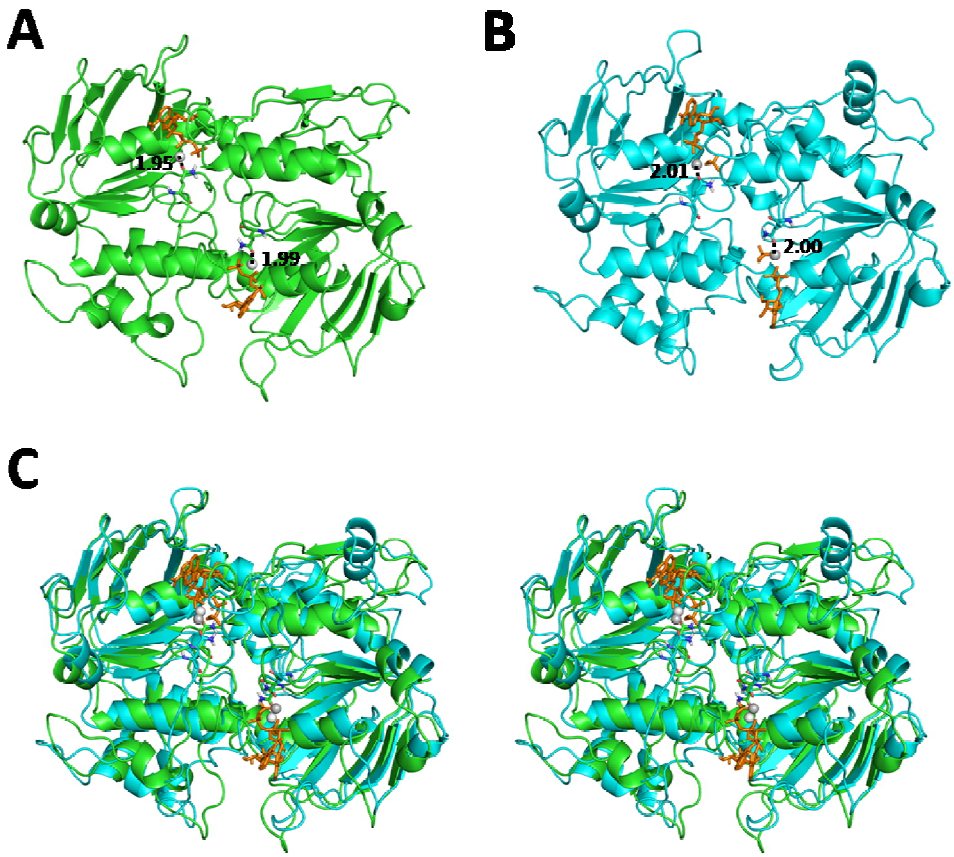


Figure 53- Helical sub-domain rotation after hydrolysis in replicate 2. The nucleotides are coloured in orange, the Q-loop glutamines are highlighted with sticks and the magnesium ions are represented as spheres. The distances between the Q-loop glutamine OE1 atom and the Mg ion are represented with a dashed line. The values for the distances are in Angstroms. Average structure for replicate 2 (over the last 10 ns of simulation after fitting to the X-ray structure) for the **A**- **2ATP** state and **B**- **2ADP.IP**

state. **C**- Stereo image representing the comparison between the ATP- and the ADP.IP-bound state average structures.

In Figure 53, which represents replicate 2, although a small rotation of the HD_1 region can be identified, no disengagement in the canonical Q-loop glutamine-Mg interaction was observed. This example clearly shows that the rotation of the HD region upon hydrolysis is independent of the state of the canonical interaction.

9.6. Relation between canonical Qloop glutamine-Mg and the NBD interface opening

In order to clarify whether there is any relation between the disengagement of the canonical Q-loop glutamine-Mg interaction and the NBD dimer opening in the **2ADP.IP** state, the distance between the RecA-like sub-domain of one NBD and the helical sub-domain of the other NBD was plotted together with distance between the Q-loop glutamine OE1 atom and the Mg atom, for replicates 4 and 5. In Figure 54A, corresponding to **2ADP.IP** replicate 5, we can observe that the dissociation of the dimer interface in site 2 is independent of the state of the canonical Q-loop glutamine-Mg interaction (which was undisturbed during the whole simulation time). A similar situation was reported by Wen and Tajkhorshid in the Maltose transporter NBD dimer opening [96]. In Figure 54B, corresponding to **2ADP.IP** replicate 4, we can see that, although the canonical interaction disappeared in the binding site 2, the NBD dimer was kept in a closed conformation, at least within the simulated timescale. From these two examples, we can conclude that dimer opening is independent of the state of the Q-loop glutamine-Mg interaction.

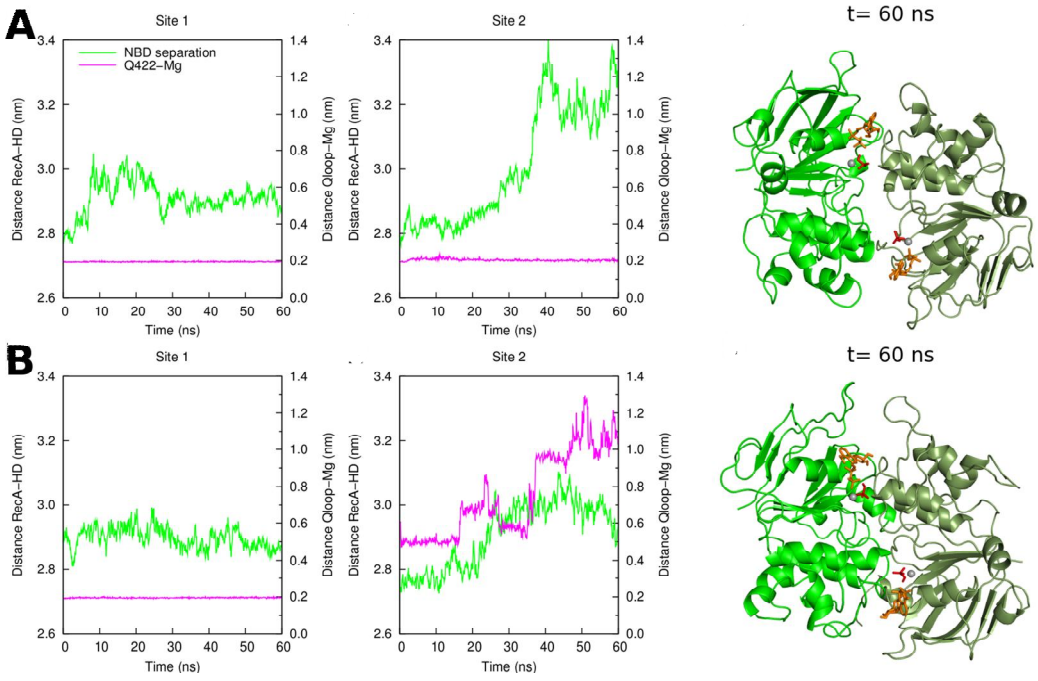


Figure 54- Relation between the dimer opening and the canonical Q-loop glutamine-Mg interaction for two **2ADP.IP** replicates (replicate 5 and 4). The time evolution of the distance between the RecA-like sub-domain of one NBD and the helical sub-domain of the other NBD is plotted in magenta in order to quantify the NBD dimer interface separation. The distance between the Q-loop glutamine OE1 atom and the Mg ion is plotted in green along the simulation time. Each point in the plots represents the distance averaged over 100 ps. **A-**Time evolution for the two described distances for **2ADP.IP** replicate 5. The right structure corresponds to the final trajectory frame (t=60 ns) for replicate 5. **B**-Same measurements as in **A** but for **2ADP.IP** replicate 4.

10. Supporting Information for chapter 5

10.1. Protonation state of protonable residues

The protonation state of each individual protonable group (residues and nucleotide species) at the pH of interest (pH=7.0), has to be specified prior to the MD simulations. This was determined using methodologies for studying the thermodynamics of proton binding, which are described elsewhere [239, 255]. These methodologies use a combination of Poisson-Boltzmann calculations, performed with the package MEAD (version 2.2.5) [234, 240, 256], and Metropolis Monte Carlo simulations, using the program PETIT (version 1.3) [239]. Since the maltose transporter is a membrane protein with an asymmetrical transmembrane pore, a custom made model was build for the calculations with this system (see Figure 55).

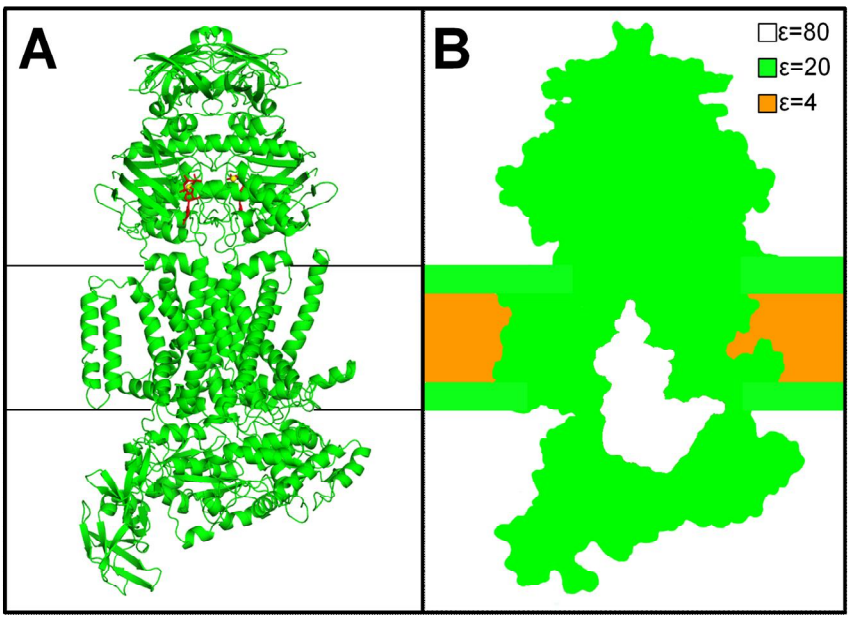


Figure 55- A schematic view of the model used to calculate the protonation states with the program MEAD (version 2.2.5) [234, 240, 256] and PETIT (version 1.3) [239] for the MalFGK$_2$E complex inserted in a lipid bilayer. **A**- Overview of the crystal structure of the complex formed by the ABC

importer MalFGK2 and the maltose binding protein (MalE). The protein is rendered as a cartoon, the ATP molecules are represented by red sticks and the magnesium ions as yellow spheres. **B**- Overview of the complete model for the MalFGK2E complex built by the package MEAD. The membrane is represented by three slabs of low dielectric media. The first and third slabs mimetize the hydrophilic lipid head groups of each leaflet of the bilayer (ε=20), and the middle slab (coloured in orange) represents the lipid hydrophobic tails (ε=4). In this model, the slabs representing the hydrophilic parts of the lipids are 7.5 Å wide, whereas the slab representing the lipid tails is 26 Å wide.

In this model, the membrane was introduced in the x-y plane as a low dielectric media formed by tree different regions. The first and third regions represents the hydrophilic lipid head groups, and is 7.5 Å wide. The second region, with a thickness of 26 Å, characterizes the middle of the bilayer and mimetizes the hydrophobic tails of the lipids. Different values of dielectric constant were assigned, based in the atomic z coordinates: for the first and third slabs of the bilayer the dielectric constant used was 20, while for the middle of the bilayer the dielectric constant used was 4. All points located inside the transmembrane pathway were treated with the dielectric constant of the solvent (80).

According to our calculations, all glutamic and aspartic acids are deprotonated with the exception of MalG Glu149. Furthermore, all lysines and arginines are protonated with the exception of MalF Lys16, Lys61 and Lys314 and MalG Lys90, Lys118 and Lys123. All free cysteines are protonated and the protonation state of all histidines is described in Table 8:

Table 8- Protonation state for all the histidine residues of the MalFGK$_2$E complex. HisA refers to the histidine residue with a zero total charge and protonated at the Nδ position, HisB corresponds also to a neutral residue but in this case protonated at Nε. The HisH represents the protonated histidine at both Nδ and Nε with the overall charge of +1.

MalG	
Residue Number	**Protonation state**
17	HisA
58	HisH
159	HisB
173	HisB

MalE	
Residue Number	**Protonation state**
39	HisH
64	HisB
203	HisH

MalK	
Residue Number	**Protonation state**
27	HisH
89	HisB
125	HisB
180	HisH
192	HisB
223	HisB
289	HisB
317	HisB
353	HisA
366	HisH

For MalK, MalF and MalG, the N- and C-terminals were considered neutral (NH_2), since the crystal structure lacked some of the first and last residues. For MalE, the N-terminal was considered positively charged (NH_3^+), whereas the C-terminal was considered negatively charged (COO^-).

10.2. DMPC bilayer construction

10.2.1. 128 DMPC bilayer

An initial model for a hydrated 128 DMPC bilayer (64 lipids in each leaflet and 4214 water molecules), obtained from Prof. Peter Tieleman's website at http://moose.bio.ucalgary.ca, was used as a starting point for the construction of our DMPC membrane. The water/lipid content was about 33 water molecules per lipid.

Since the initial membrane model had been equilibrated under different simulation conditions, we submitted the 128 DMPC molecules for 30 ns. The simulation was performed using the GROMACS 3.3.1 package [213, 214, 257, 307]. The lipid parameters used were based on the 53A6 GROMOS96 force field [196, 197, 199, 200], except for the atomic partial charges, which had been previously derived by Chiu and coworkers [275]. The membrane equilibration was performed at constant temperature conditions (310K). This temperature is above the phase transition temperature reported in literature for DMPC (Tm=296-297K) in order to ensure that the bilayer is in the liquid crystalline state [276]. The setup for the molecular dynamics simulations is similar to the one reported in section 5 in the Materials and Methods.

The evolution for some lipid properties (area and volume per lipid) along the simulation time can be followed in Figure 56. As can be seen, in the end of the 30 ns of simulation, the area (Figure 56A) and volume per lipid (Figure 56B), despite showing some fluctuations, remain more or less within the experimental range. Based on previous works with DMPC lipidic bilayers [310], we considered the first 25 ns as an equilibration period, and this data was discarded. The average values for the area and volume per lipid over the last 5 ns of simulation are 0.648 nm^2 and 1.098 nm^3, respectively. Additionally, the organization of the hydrocarbon tails can be calculated from the simulation and compared to experimental results obtained using NMR techniques (Figure 56C). As can be seen in Figure 56C, the values for the deuterium order parameters from the simulations are in good agreement with the values obtained experimentally [309].

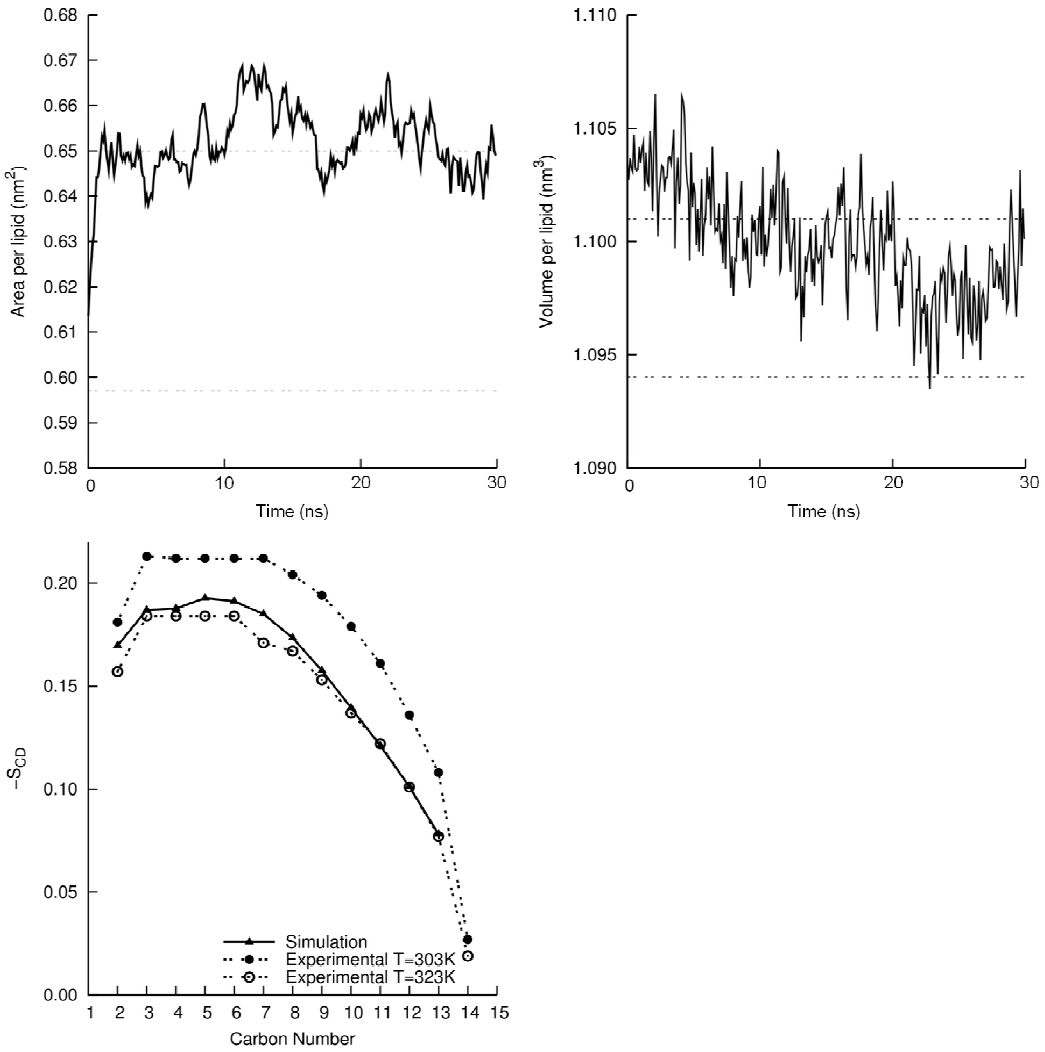


Figure 56- **A-** Area per lipid for the 128 DMPC as a function of the simulation time (solid black line). The horizontal dashed lines represent the upper and lower limit ranges for the experimental values [308]. **B**- Time evolution of the volume per lipid (solid black line). Again, the horizontal dashed lines represent the upper and lower limit ranges for the experimental values [308]. In both plots, each point represents the average value over 100 ps. **C**- Deuterium order parameters. Black and white circles are the experimental results [309] averaged over both sn-1 and sn-2 chains for two different temperatures (T=303K and T=323K), while the black triangles represent the data obtained from the present simulation.

10.2.2. 512 DMPC bilayer

The 512DMPC bilayer was generated by the replication of the already described pre-equilibrated 128 DMPC bilayer, in both the x and y directions. The resulting system was then energy minimized and re-equilibrated again for over 20 ns. The system was subjected to 5000 steps of steepest descent energy minimization in order to remove excessive strain, and then submitted to 100 ps of restrained MD simulation: the first 20 ps with restraints in all heavy atoms, including water (force constant 1000 kJ mol^{-1} nm^{-2}), followed by 30 ps with restraints in all lipid heavy atoms (force constant 1000 kJ mol^{-1} nm^{-2}) and the last 50 ps with restraints in all the lipid heavy atoms, but in this case with a lower value for the force restraints (5 kJ mol^{-1} nm^{-2}). The unrestrained MD simulations started from this point forward and lasted 20 ns. All the simulation details were exactly like the ones described in the main article in the Materials and Methods section. In the end of the re-equilibration period, similarly to the 128 DMPC bilayer, the 512DMPC bilayer presents area and volume per lipid values within the experimental range. The average value for the area and volume per lipid over the last 5 ns of simulation is 0.658 nm^2 and 1.099 nm^3, respectively.

10.3. MalFGK$_2$E behaviour during the MD simulations

For all states (**2ATP, 2ADP.IP** and **2ADP**), the protein behaviour during the 50 ns of simulation was carefully examined for all ten replicates, by the determination of the root mean square deviation from the X-ray structure (see Figure 38) and the secondary structure content (Figure 57). In general, the RMSD values reported for our three states, although large, are in the same order of magnitude as the ones reported in previous MD studies for other ABC transporters [162, 164, 168]. Additionally, the percentage of secondary structure content (averaged over the ten replicates) was determined over the simulation time in order to see if abnormal structure loss was observable (Figure 57). As can be seen from Figure 57, the overall secondary structure pattern is maintained, for both the pre- and post-

hydrolysis simulated states, and only a small (<12%) secondary structure loss was measured. The RMSD values, together with the time evolution of native secondary structure content, are indicative of structural stability for the MalFGK$_2$E complex.

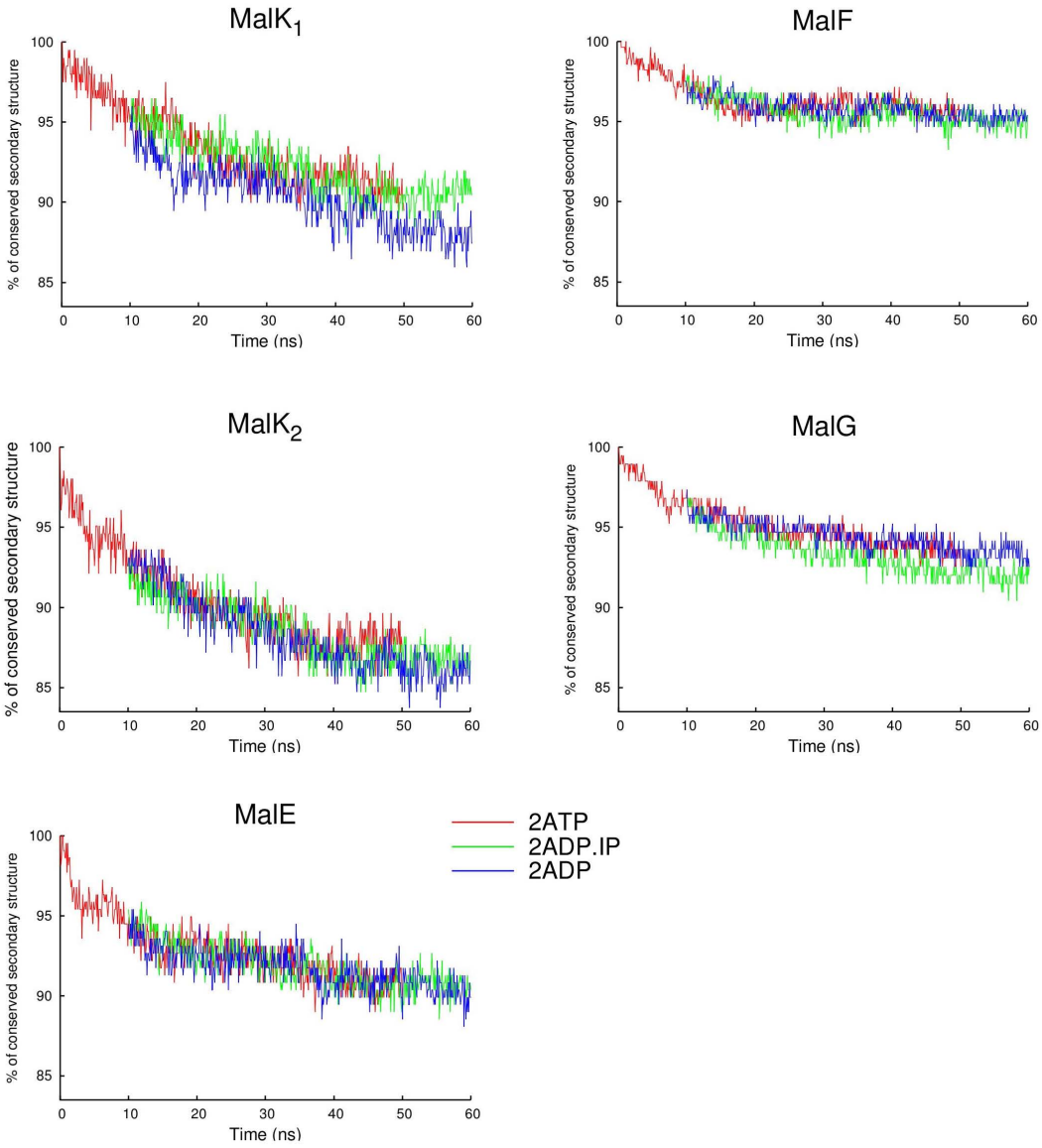


Figure 57- Percentage of residues within the same secondary structure class as in the X-ray structure (calculated using DSSP [262]) for MalK$_1$, MalK$_2$, MalF, MalG and MalE for the **2ATP** (red line),

2ADP.IP (green line) and the **2ADP** (blue line) states. The averages were obtained over all 10 replicates for each state. The regular secondary structures considered for this measure were the α-helix, the β-sheet, the 3_{10} helix and the β-bridge (DSSP classification [262]).

10.4. Rigid body motion of the MalF Periplasmic region

A large amplitude rigid body rotation of the second periplasmic region of MalF (MalF-P_2 region) toward the membrane was observed in the **2ATP** state (see Figure 58). As can be seen from Figure 58, the MalF-P_2 region rotation was detected for all ten replicates, and it takes 5 to 20 ns to be completed. Additionally, for the **2ADP.IP** and **2ADP** states, a high amplitude rigid body rotation (similar to the one reported for the **2ATP** state) was also observable for all replicates (see Figure 58B and Figure 58C).

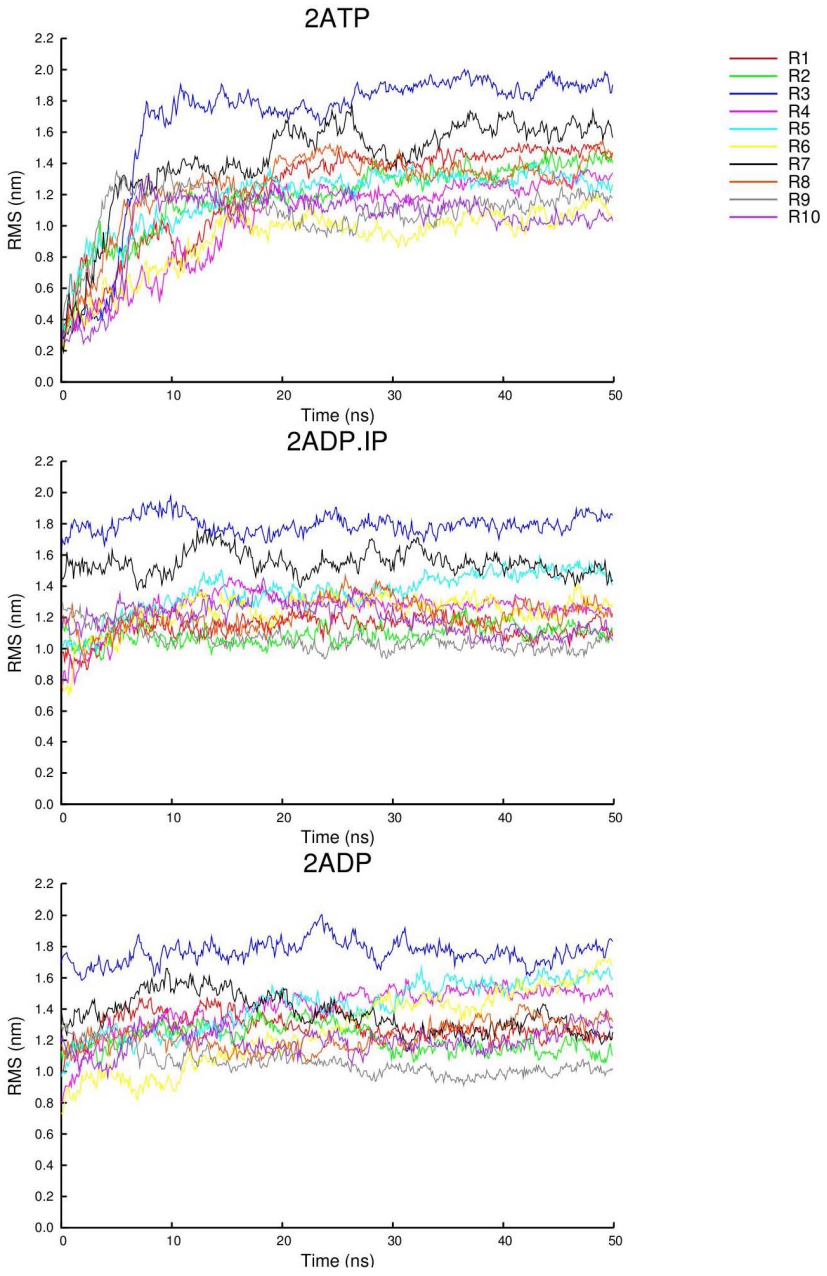


Figure 58- Time evolution of the distance between the C_α atoms of the MalF-P_2 region and its position in the crystal structure, after fitting to the Cα MalF core transmembrane region, for the **2ATP**, **2ADP.IP** and the **2ADP** states. Each point represents the average value over 100 ps.

10.5. MalK dimer interface

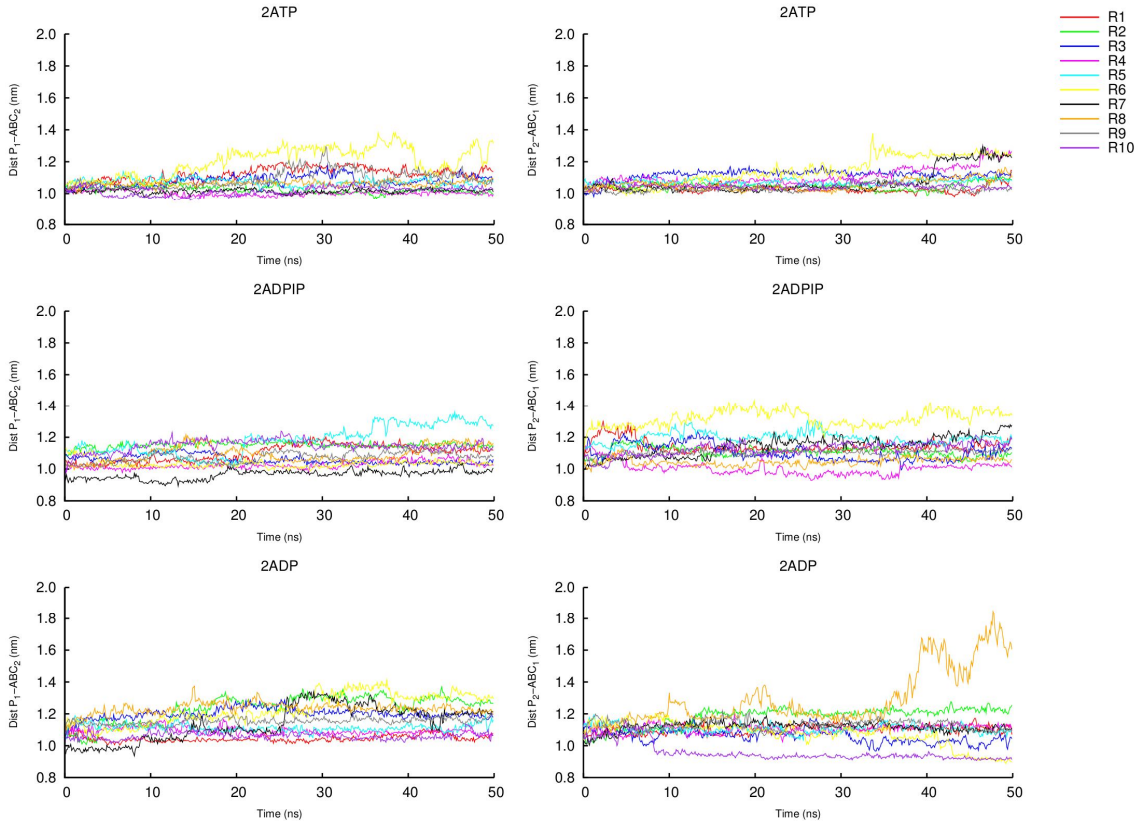


Figure 59- Distances, for all replicates, between the P-loop motif of one MalK monomer and the ABC-signature motif of the other. The graphics at the right side represent the distances for binding site 1, whereas the ones at the left represent the same measurement for binding site 2. Each point represents the average value over 100 ps.

In order to determine if the MalK dimer dissociation was observable in any of the **2ATP**, **2ADP.IP** and **2ADP** replicates, the distance between the P-loop of one MalK monomer and the ABC signature motif of the other was determined. As can be seen in Figure 59, no MalK dimer interface separation was detected in the **2ATP** and **2ADP.IP** states. In contrast, one of the ten **2ADP** replicates (replicate 8) clearly

shows, after 37 ns of simulation, an interface opening in the nucleotide binding site 2 only.

The effect of the nucleotide binding site 2 dissociation in the MalF and MalG domains is small, and no significant conformational rearrangements in the transmembrane domains were detected, as can be confirmed in Figure 60.

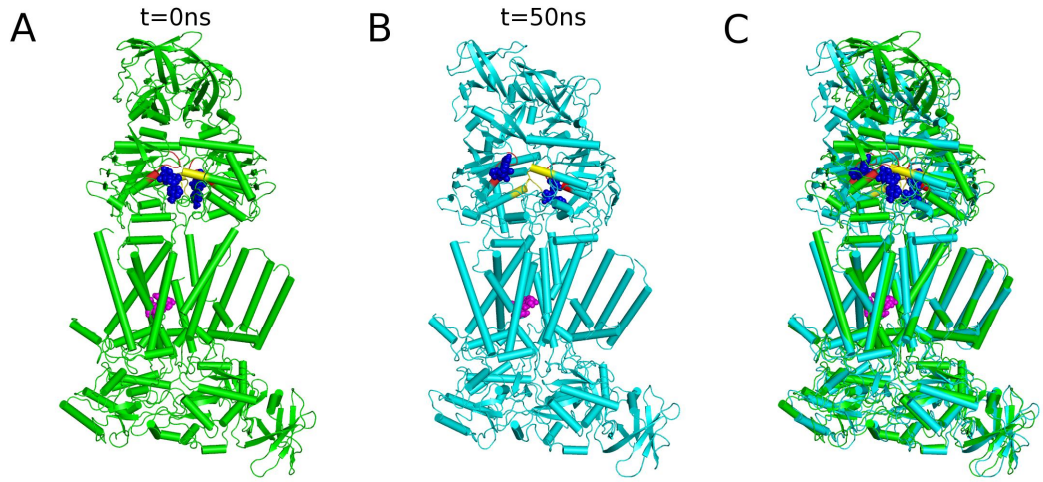


Figure 60- Opening of the MalK dimer interface and the absence of significant effects on the transmembrane domains. Snapshots of the **2ADP** replicate 8 simulation (the only replicate exhibiting a MalK dimer dissociation). The P-loop and the ABC motif are colored in red and yellow, respectively, whereas ADP and maltose are represented by dark blue and magenta spheres. **A**) Starting point (t=0ns). **B**) Final structure (after 50 ns). **C**) Superimposition of the starting and the final structures after fitting to the core transmembrane helices (residues 13-90 and 272-504 from MalF and 7-296 from MalG).

10.6. Helical sub-domain rotation

The helical sub-domain (HD) region has been identified in the last years as one of the regions presenting significant conformational changes during the ATP-cycle [1,

35, 36, 51-53, 87], and is most probably involved in the NBD-TMD communication mechanism [297]. The rigid body outward rotation of the HD is thought to constitute one of the major structural differences between different nucleotide bound states in the NBDs of ABC transporters. In our simulations, in both post-hydrolysis states, a small amplitude rotation of the HD, relative to the pre-hydrolysis state, was observable (see Figure 61).

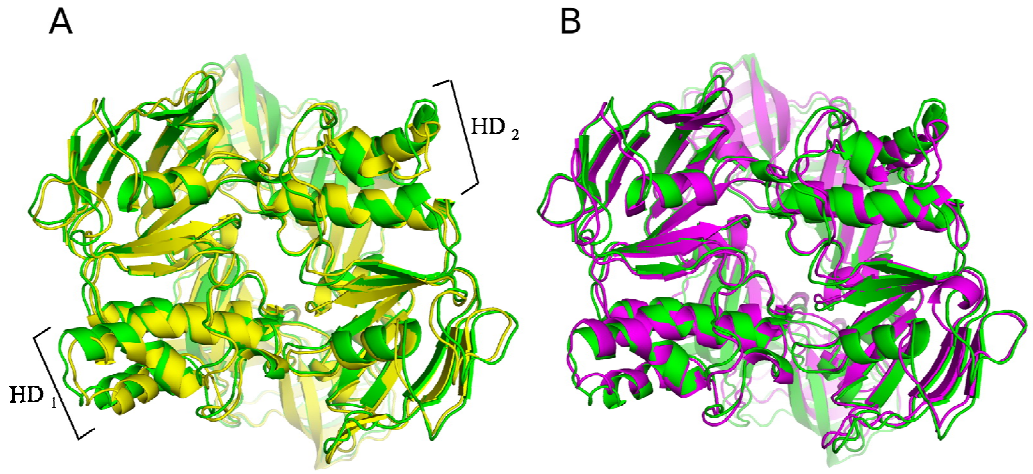


Figure 61- Helical sub-domain rotation upon hydrolysis and exit of the IP species. Comparison between the average structures (obtained over the last 10 ns of simulation) for the **2ATP** (colored in green) and the: **A**- **2ADP.IP** state (colored in yellow). **B**- **2ADP** state (colored in magenta). These images were obtained after superimposition of the RecA-like sub-domains (residues 4 to 87 and 151 to 216) between the pre- and post-hydrolysis states. The regulatory domains are shown in the back with faded colors. For simplicity purposes, the MalF, MalG, MalE and the nucleotide species were not represented.

10.7. The position of maltose during ATP hydrolysis

In order to clarify if the ATP-hydrolytic cycle exerts any effect in the allocrite binding site, the maltose position, relative to its starting position, was determined, along the simulation time, for all replicates of all states.

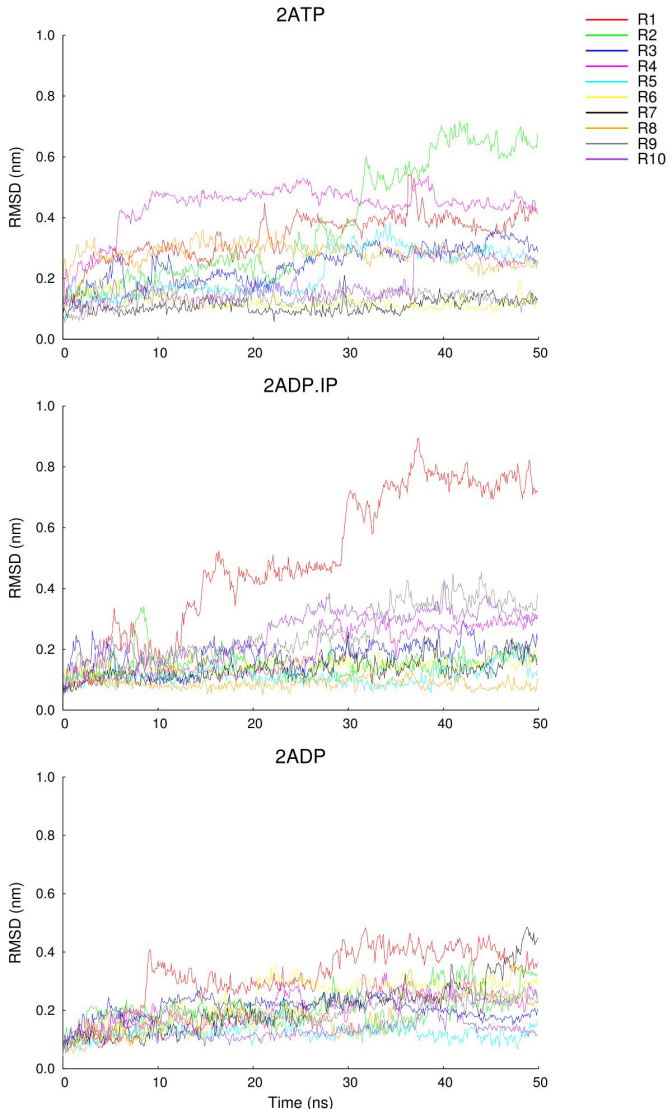


Figure 62- Time evolution of the position of maltose relative to its position in the starting structure, after root mean square fitting to the core transmembrane domains (MalF residue 13-90 and 272-504 and MalG residue 7-296) Cα atoms. Each point represents the average value over 100 ps

As can be seen in Figure 62, the maltose molecule does not abandon the allocrite binding site in the majority of the simulations. However, two exceptions can be detected: one in the **2ATP** state (replicate 2) and the other in the **2ADP.IP** state (replicate 1). Moreover, a detailed view of the maltose movement from the allocrite binding site can be observed in Figure 63.

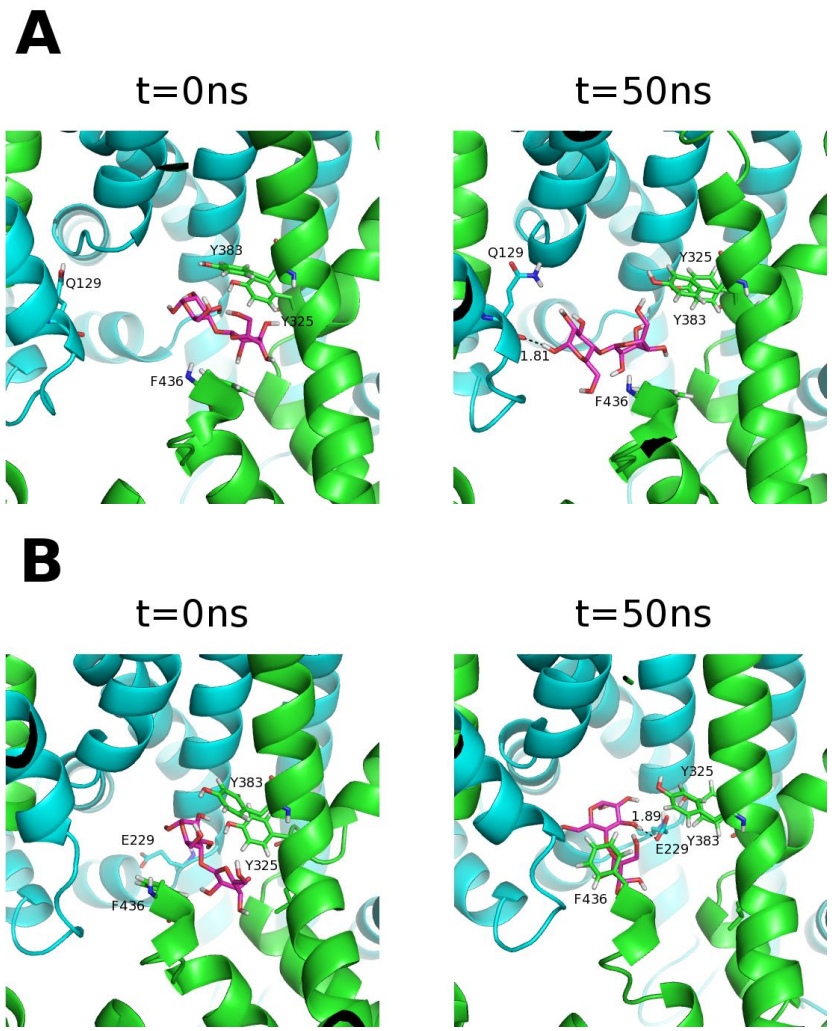


Figure 63- Snapshots of the two cases (replicate 2 of the **2ATP** state and replicate 1 of the **2ADP.IP** state) where maltose release from the allocrite binding site is observed. In these figures, MalF is

colored in green, whereas MalG is colored in cyan. Some of the MalF (binding site) and MalG residues are represented in green and cyan sticks. The distance between maltose and the MalG residues is shown with a dashed line (Å). Snapshots of the allocrite binding site at the beginning (t=0 ns) and at the end (t=50 ns) of the simulation for the **A**) **2ATP** replicate 2. **B**) **2ADP.IP** replicate 1.

The allocrite binding site region is not affected by ATP-hydrolysis or IP exit, since no significant differences in the binding site region were observed in the three studied states, at least within the simulated timescales (Figure 64).

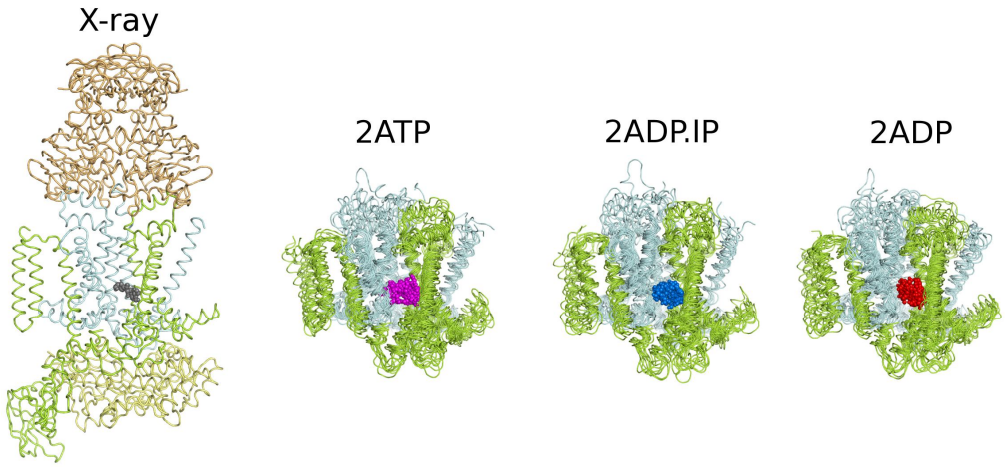


Figure 64- Maltose position in the transmembrane domains. In this figure, MalF is colored in green, whereas MalG is highlighted in blue. The left side image represents the X-ray structure with the MalK dimer, colored in light orange, and MalE colored in brown. The maltose molecule is represented with gray spheres. The remaining three images represent the last conformations (after 50 ns) for each one of the 10 replicates of the three simulated states, after fitting to the transmembrane core (MalF residues 13-90 and 272-504 and MalG 7-296). In the **2ATP, 2ADP.IP** and **2ADP,** only the MalF and MalG transmembrane regions are represented, for simplicity reasons, and the maltose molecule is shown in magenta, blue and red, respectively.

11. Bibliography

1. Oliveira, A.S., A.M. Baptista, and C.M. Soares, *Insights into the molecular mechanism of an ABC transporter: conformational changes in the NBD dimer of MJ0796.* J Phys Chem B, 2010. **114**(16): p. 5486-96.
2. Oliveira, A.S., A.M. Baptista, and C.M. Soares, *Conformational changes induced by ATP-hydrolysis in an ABC transporter: a molecular dynamics study of the Sav1866 exporter.* submitted to J Phys Chem B, 2010.
3. Oliveira, A.S., A.M. Baptista, and C.M. Soares, *Inter-domain communication mechanisms in an ABC importer: A molecular dynamics study of the MalFGK2E complex.* manuscript in preparation, 2010.
4. Lodish, H., et al., *Molecular Cell Biology*. Third Edition ed. 1998, New York: Scientific American Books.
5. Higgins, C.F., *ABC transporters: from microorganisms to man.* Annu Rev Cell Biol, 1992. **8**: p. 67-113.
6. Oswald, C., I.B. Holland, and L. Schmitt, *The motor domains of ABC-transporters. What can structures tell us?* Naunyn Schmiedebergs Arch Pharmacol, 2006. **372**(6): p. 385-99.
7. Davidson, A.L., et al., *Structure, function, and evolution of bacterial ATP-binding cassette systems.* Microbiol Mol Biol Rev, 2008. **72**(2): p. 317-64.
8. Higgins, C.F., et al., *A family of related ATP-binding subunits coupled to many distinct biological processes in bacteria.* Nature, 1986. **323**(6087): p. 448-50.
9. Hyde, S.C., et al., *Structural Model of ATP-Binding Proteins Associated with Cystic-Fibrosis, Multidrug Resistance and Bacterial Transport.* Nature, 1990. **346**(6282): p. 362-365.
10. Goosen, N. and G.F. Moolenaar, *Role of ATP hydrolysis by UvrA and UvrB during nucleotide excision repair.* Res Microbiol, 2001. **152**(3-4): p. 401-9.
11. Chakraburtty, K., *Translational regulation by ABC systems.* Res Microbiol, 2001. **152**(3-4): p. 391-9.
12. Hirano, T., *At the heart of the chromosome: SMC proteins in action.* Nat Rev Mol Cell Biol, 2006. **7**(5): p. 311-22.
13. Doolittle, R.F., et al., *Domainal Evolution of a Prokaryotic DNA-Repair Protein and Its Relationship to Active-Transport Proteins.* Nature, 1986. **323**(6087): p. 451-453.
14. Locher, K.P., *Review. Structure and mechanism of ATP-binding cassette transporters.* Philos Trans R Soc Lond B, 2009. **364**(1514): p. 239-45.
15. Borst, P. and R.O. Elferink, *Mammalian ABC transporters in health and disease.* Annu Rev Biochem, 2002. **71**: p. 537-92.
16. Seeger, M.A. and H.W. van Veen, *Molecular basis of multidrug transport by ABC transporters.* Biochim Biophys Acta, 2009. **1794**(5): p. 725-37.

17. Dean, M., Y. Hamon, and G. Chimini, *The human ATP-binding cassette (ABC) transporter superfamily.* J Lipid Res, 2001. **42**(7): p. 1007-1017.
18. Albrecht, C., et al., *A novel missense mutation in ABCA1 results in altered protein trafficking and reduced phosphatidylserine translocation in a patient with Scott syndrome.* Blood, 2005. **106**(2): p. 542-9.
19. Martinez-Mir, A., et al., *Retinitis pigmentosa caused by a homozygous mutation in the Stargardt disease gene ABCR.* Nat Genet, 1998. **18**(1): p. 11-2.
20. Aguirre-Lamban, J., et al., *Novel human pathological mutations. Gene symbol: ABCA4. Disease: Stargardt disease.* Hum Genet, 2010. **127**(1): p. 119.
21. Jacquemin, E., *Progressive familial intrahepatic cholestasis. Genetic basis and treatment.* Clin Liver Dis, 2000. **4**(4): p. 753-63.
22. Riordan, J.R., et al., *Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA.* Science, 1989. **245**(4922): p. 1066-73.
23. Robey, R.W., et al., *Overexpression of the ATP-binding cassette half-transporter, ABCG2 (Mxr/BCrp/ABCP1), in flavopiridol-resistant human breast cancer cells.* Clin Cancer Res, 2001. **7**(1): p. 145-52.
24. Nikaido, H., *Multidrug resistance in bacteria.* Annu Rev Biochem, 2009. **78**: p. 119-46.
25. Cannon, R.D., et al., *Efflux-mediated antifungal drug resistance.* Clin Microbiol Rev, 2009. **22**(2): p. 291-321.
26. Higgins, C.F., et al., *Complete nucleotide sequence and identification of membrane components of the histidine transport operon of S. typhimurium.* Nature, 1982. **298**(5876): p. 723-7.
27. Holland, I.B. and M.A. Blight, *ABC-ATPases, adaptable energy generators fuelling transmembrane movement of a variety of molecules in organisms from bacteria to humans.* J Mol Biol, 1999. **293**(2): p. 381-99.
28. Biemans-Oldehinkel, E., M.K. Doeven, and B. Poolman, *ABC transporter architecture and regulatory roles of accessory domains.* FEBS Lett, 2006. **580**(4): p. 1023-35.
29. Cheng, S.H., et al., *Phosphorylation of the R domain by cAMP-dependent protein kinase regulates the CFTR chloride channel.* Cell, 1991. **66**(5): p. 1027-36.
30. Rich, D.P., et al., *Effect of deleting the R domain on CFTR-generated chloride channels.* Science, 1991. **253**(5016): p. 205-7.
31. Reidl, J., et al., *MalI, a novel protein involved in regulation of the maltose system of Escherichia coli, is highly homologous to the repressor proteins GalR, CytR, and LacI.* J Bacteriol, 1989. **171**(9): p. 4888-99.
32. Ames, G.F., *Bacterial periplasmic transport systems: structure, mechanism, and evolution.* Annu Rev Biochem, 1986. **55**: p. 397-425.

33. Hung, L.W., et al., *Crystal structure of the ATP-binding subunit of an ABC transporter.* Nature, 1998. **396**(6712): p. 703-7.
34. Gaudet, R. and D.C. Wiley, *Structure of the ABC ATPase domain of human TAP1, the transporter associated with antigen processing.* Embo J, 2001. **20**(17): p. 4964-72.
35. Karpowich, N., et al., *Crystal structures of the MJ1267 ATP binding cassette reveal an induced-fit effect at the ATPase active site of an ABC transporter.* Structure (Camb), 2001. **9**(7): p. 571-86.
36. Yuan, Y.R., et al., *The crystal structure of the MJ0796 ATP-binding cassette. Implications for the structural consequences of ATP hydrolysis in the active site of an ABC transporter.* J Biol Chem, 2001. **276**(34): p. 32313-21.
37. Schmitt, L., et al., *Crystal structure of the nucleotide-binding domain of the ABC-transporter haemolysin B: identification of a variable region within ABC helical domains.* J Mol Biol, 2003. **330**(2): p. 333-42.
38. Verdon, G., et al., *Crystal structures of the ATPase subunit of the glucose ABC transporter from Sulfolobus solfataricus: nucleotide-free and nucleotide-bound conformations.* J Mol Biol, 2003. **330**(2): p. 343-58.
39. Verdon, G., et al., *Formation of the productive ATP-Mg2+-bound dimer of GlcV, an ABC-ATPase from Sulfolobus solfataricus.* J Mol Biol, 2003. **334**(2): p. 255-67.
40. Lewis, H.A., et al., *Structure of nucleotide-binding domain 1 of the cystic fibrosis transmembrane conductance regulator.* Embo J, 2004. **23**(2): p. 282-93.
41. Ose, T., et al., *Crystal structure of the ATP-binding cassette of multisugar transporter from Pyrococcus horikoshii OT3.* Proteins, 2004. **57**(3): p. 635-8.
42. Lewis, H.A., et al., *Impact of the deltaF508 mutation in first nucleotide-binding domain of human cystic fibrosis transmembrane conductance regulator on domain folding and structure.* J Biol Chem, 2005. **280**(2): p. 1346-53.
43. Scheffel, F., et al., *Structure of the ATPase subunit CysA of the putative sulfate ATP-binding cassette (ABC) transporter from Alicyclobacillus acidocaldarius.* FEBS Lett, 2005. **579**(13): p. 2953-8.
44. Thibodeau, P.H., et al., *Side chain and backbone contributions of Phe508 to CFTR folding.* Nat Struct Mol Biol, 2005. **12**(1): p. 10-6.
45. Watanabe, S., A. Kita, and K. Miki, *Crystal structure of atypical cytoplasmic ABC-ATPase SufC from Thermus thermophilus HB8.* J Mol Biol, 2005. **353**(5): p. 1043-54.
46. Procko, E., et al., *Distinct structural and functional properties of the ATPase sites in an asymmetric ABC transporter.* Mol Cell, 2006. **24**(1): p. 51-62.
47. Zaitseva, J., et al., *A structural analysis of asymmetry required for catalytic activity of an ABC-ATPase domain dimer.* Embo J, 2006. **25**(14): p. 3432-43.

48. Ramaen, O., et al., *Structure of the human multidrug resistance protein 1 nucleotide binding domain 1 bound to Mg2+/ATP reveals a non-productive catalytic site.* J Mol Biol, 2006. **359**(4): p. 940-9.
49. Vedadi, M., et al., *Genome-scale protein expression and structural biology of Plasmodium falciparum and related Apicomplexan organisms.* Mol Biochem Parasitol, 2007. **151**(1): p. 100-10.
50. Diederichs, K., et al., *Crystal structure of MalK, the ATPase subunit of the trehalose/maltose ABC transporter of the archaeon Thermococcus litoralis.* Embo J, 2000. **19**(22): p. 5951-61.
51. Smith, P.C., et al., *ATP binding to the motor domain from an ABC transporter drives formation of a nucleotide sandwich dimer.* Mol Cell, 2002. **10**(1): p. 139-49.
52. Chen, J., et al., *A tweezers-like motion of the ATP-binding cassette dimer in an ABC transport cycle.* Mol Cell, 2003. **12**(3): p. 651-61.
53. Lu, G., et al., *ATP hydrolysis is required to reset the ATP-binding cassette dimer into the resting-state conformation.* Proc Natl Acad Sci U S A, 2005. **102**(50): p. 17969-74.
54. Zaitseva, J., et al., *H662 is the linchpin of ATP hydrolysis in the nucleotide-binding domain of the ABC transporter HlyB.* Embo J, 2005. **24**(11): p. 1901-10.
55. Kitaoka, S., et al., *Crystal structure of Escherichia coli SufC, an ABC-type ATPase component of the SUF iron-sulfur cluster assembly machinery.* FEBS Lett, 2006. **580**(1): p. 137-43.
56. Newstead, S., et al., *Insights into how nucleotide-binding domains power ABC transport.* Structure, 2009. **17**(9): p. 1213-22.
57. Locher, K.P., A.T. Lee, and D.C. Rees, *The E. coli BtuCD structure: a framework for ABC transporter architecture and mechanism.* Science, 2002. **296**(5570): p. 1091-8.
58. Dawson, R.J. and K.P. Locher, *Structure of a bacterial multidrug ABC transporter.* Nature, 2006. **443**(7108): p. 180-5.
59. Dawson, R.J. and K.P. Locher, *Structure of the multidrug ABC transporter Sav1866 from Staphylococcus aureus in complex with AMP-PNP.* FEBS Lett, 2007. **581**(5): p. 935-8.
60. Hollenstein, K., D.C. Frei, and K.P. Locher, *Structure of an ABC transporter in complex with its binding protein.* Nature, 2007. **446**(7132): p. 213-6.
61. Hvorup, R.N., et al., *Asymmetry in the structure of the ABC transporter-binding protein complex BtuCD-BtuF.* Science, 2007. **317**(5843): p. 1387-90.
62. Oldham, M.L., et al., *Crystal structure of a catalytic intermediate of the maltose transporter.* Nature, 2007. **450**(7169): p. 515-U7.
63. Pinkett, H.W., et al., *An inward-facing conformation of a putative metal-chelate-type ABC transporter.* Science, 2007. **315**(5810): p. 373-7.
64. Ward, A., et al., *Flexibility in the ABC transporter MsbA: Alternating access with a twist.* Proc Natl Acad Sci U S A, 2007. **104**(48): p. 19005-10.

65. Aller, S.G., et al., *Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding.* Science, 2009. **323**(5922): p. 1718-22.
66. Kadaba, N.S., et al., *The high-affinity E. coli methionine ABC transporter: structure and allosteric regulation.* Science, 2008. **321**(5886): p. 250-3.
67. Dassa, E. and M. Hofnung, *Sequence of gene malG in E. coli K12: homologies between integral membrane components from binding protein-dependent transport systems.* Embo J, 1985. **4**(9): p. 2287-93.
68. Saurin, W., W. Koster, and E. Dassa, *Bacterial binding protein-dependent permeases: characterization of distinctive signatures for functionally related integral cytoplasmic membrane proteins.* Mol Microbiol, 1994. **12**(6): p. 993-1004.
69. Pearce, S.R., et al., *Membrane topology of the integral membrane components, OppB and OppC, of the oligopeptide permease of Salmonella typhimurium.* Mol Microbiol, 1992. **6**(1): p. 47-57.
70. Hollenstein, K., R.J. Dawson, and K.P. Locher, *Structure and mechanism of ABC transporter proteins.* Curr Opin Struct Biol, 2007.
71. Rees, D.C., E. Johnson, and O. Lewinson, *ABC transporters: the power to change.* Nat Rev Mol Cell Biol, 2009. **10**(3): p. 218-27.
72. Davidson, A.L. and J. Chen, *ATP-binding cassette transporters in bacteria.* Annu Rev Biochem, 2004. **73**: p. 241-68.
73. Walker, J.E., et al., *Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold.* Embo J, 1982. **1**(8): p. 945-51.
74. Higgins, C.F., et al., *Nucleotide binding by membrane components of bacterial periplasmic binding protein-dependent transport systems.* Embo J, 1985. **4**(4): p. 1033-9.
75. Shyamala, V., et al., *Structure-function analysis of the histidine permease and comparison with cystic fibrosis mutations.* J Biol Chem, 1991. **266**(28): p. 18714-9.
76. Moussatova, A., et al., *ATP-binding cassette transporters in Escherichia coli.* Biochim Biophys Acta, 2008. **1778**(9): p. 1757-71.
77. Ambudkar, S.V., et al., *The A-loop, a novel conserved aromatic acid subdomain upstream of the Walker A motif in ABC transporters, is critical for ATP binding.* FEBS Lett, 2006. **580**(4): p. 1049-55.
78. Hopfner, K.P., et al., *Structural biology of Rad50 ATPase: ATP-driven conformational control in DNA double-strand break repair and the ABC-ATPase superfamily.* Cell, 2000. **101**(7): p. 789-800.
79. Geourjon, C., et al., *A common mechanism for ATP hydrolysis in ABC transporter and helicase superfamilies.* Trends Biochem Sci, 2001. **26**(9): p. 539-44.

80. Moody, J.E., et al., *Cooperative, ATP-dependent association of the nucleotide binding cassettes during the catalytic cycle of ATP-binding cassette transporters.* J Biol Chem, 2002. **277**(24): p. 21111-4.
81. Orelle, C., et al., *The conserved glutamate residue adjacent to the Walker-B motif is the catalytic base for ATP hydrolysis in the ATP-binding cassette transporter BmrA.* J Biol Chem, 2003. **278**(47): p. 47002-8.
82. Tombline, G., et al., *Properties of P-glycoprotein with mutations in the "catalytic carboxylate" glutamate residues.* J Biol Chem, 2004. **279**(45): p. 46518-26.
83. Urbatsch, I.L., et al., *Mutational analysis of conserved carboxylate residues in the nucleotide binding sites of P-glycoprotein.* Biochemistry, 2000. **39**(46): p. 14138-49.
84. Ernst, R., et al., *Engineering ATPase activity in the isolated ABC cassette of human TAP1.* J Biol Chem, 2006. **281**(37): p. 27471-80.
85. Davidson, A.L. and S. Sharma, *Mutation of a single MalK subunit severely impairs maltose transport activity in Escherichia coli.* J Bacteriol, 1997. **179**(17): p. 5458-64.
86. Jones, P.M. and A.M. George, *The ABC transporter structure and mechanism: perspectives on recent research.* Cell Mol Life Sci, 2004. **61**(6): p. 682-99.
87. Jones, P.M. and A.M. George, *Nucleotide-dependent allostery within the ABC transporter ATP-binding cassette: a computational study of the MJ0796 dimer.* J Biol Chem, 2007. **282**(31): p. 22793-803.
88. Davidson, A.L., S.S. Laghaeian, and D.E. Mannering, *The maltose transport system of Escherichia coli displays positive cooperativity in ATP hydrolysis.* J Biol Chem, 1996. **271**(9): p. 4858-63.
89. Steinfels, E., et al., *Characterization of YvcC (BmrA), a multidrug ABC transporter constitutively expressed in Bacillus subtilis.* Biochemistry, 2004. **43**(23): p. 7491-502.
90. Zaitseva, J., et al., *Functional characterization and ATP-induced dimerization of the isolated ABC-domain of the haemolysin B transporter.* Biochemistry, 2005. **44**(28): p. 9680-90.
91. Karpowich, N.K., et al., *Crystal structures of the BtuF periplasmic-binding protein for vitamin B12 suggest a functionally important reduction in protein mobility upon ligand binding.* J Biol Chem, 2003. **278**(10): p. 8429-34.
92. Balan, A., et al., *Crystallographic structure and substrate-binding interactions of the molybdate-binding protein of the phytopathogen Xanthomonas axonopodis pv. citri.* Biochim Biophys Acta-Proteins and Proteomics, 2008. **1784**(2): p. 393-399.
93. Oloo, E.O. and D.P. Tieleman, *Conformational transitions induced by the binding of MgATP to the vitamin B12 ATP-binding cassette (ABC) transporter BtuCD.* J Biol Chem, 2004. **279**(43): p. 45013-9.

94. Oloo, E.O., E.Y. Fung, and D.P. Tieleman, *The dynamics of the MgATP-driven closure of MalK, the energy-transducing subunit of the maltose ABC transporter.* J Biol Chem, 2006. **281**(38): p. 28397-407.

95. Ivetac, A., J.D. Campbell, and M.S. Sansom, *Dynamics and function in a bacterial ABC transporter: simulation studies of the BtuCDF system and its components.* Biochemistry, 2007. **46**(10): p. 2767-78.

96. Wen, P.C. and E. Tajkhorshid, *Dimer opening of the nucleotide binding domains of ABC transporters after ATP hydrolysis.* Biophys J, 2008. **95**(11): p. 5100-10.

97. Jones, P.M. and A.M. George, *Opening of the ADP-bound active site in the ABC transporter ATPase dimer: evidence for a constant contact, alternating sites model for the catalytic cycle.* Proteins, 2009. **75**(2): p. 387-96.

98. Damas, J.M., et al., *Structural consequences of ATP hydrolysis on the ABC transporter NBD dimer: molecular dynamics studies of HlyB.* Manuscript submitted to Proteins, 2010.

99. Kreimer, D.I., K.P. Chai, and G. Ferro-Luzzi Ames, *Nonequivalence of the nucleotide-binding subunits of an ABC transporter, the histidine permease, and conformational changes in the membrane complex.* Biochemistry, 2000. **39**(46): p. 14183-95.

100. Hunke, S., et al., *ATP modulates subunit-subunit interactions in an ATP-binding cassette transporter (MalFGK2) determined by site-directed chemical cross-linking.* J Biol Chem, 2000. **275**(20): p. 15526-34.

101. Daus, M.L., et al., *ATP-driven MalK dimer closure and reopening and conformational changes of the "EAA" motifs are crucial for function of the maltose ATP-binding cassette transporter (MalFGK2).* J Biol Chem, 2007. **282**(31): p. 22387-96.

102. Quiocho, F.A., *Atomic structures of periplasmic binding proteins and the high-affinity active transport systems in bacteria.* Philos Trans R Soc Lond B Biol Sci, 1990. **326**(1236): p. 341-51; discussion 351-2.

103. Urbatsch, I.L., et al., *P-Glycoprotein Is Stably Inhibited by Vanadate-Induced Trapping of Nucleotide at a Single Catalytic Site.* J Biol Chem, 1995. **270**(33): p. 19383-19390.

104. Sharma, S. and A.L. Davidson, *Vanadate-induced trapping of nucleotides by purified maltose transport complex requires ATP hydrolysis.* J Bacteriol, 2000. **182**(23): p. 6570-6.

105. Siarheyeva, A., R. Liu, and F.J. Sharom, *Characterization of an asymmetric occluded state of P-glycoprotein with two bound nucleotides: implications for catalysis.* J Biol Chem, 2010. **285**(10): p. 7575-86.

106. Liu, P.Q., C.E. Liu, and G.F. Ames, *Modulation of ATPase activity by physical disengagement of the ATP-binding domains of an ABC transporter, the histidine permease.* J Biol Chem, 1999. **274**(26): p. 18310-8.

107. Buckel, S.D., et al., *An analysis of the structure of the product of the rbsA gene of Escherichia coli K12.* J Biol Chem, 1986. **261**(17): p. 7659-62.

108. Nikaido, K. and G.F. Ames, *One intact ATP-binding subunit is sufficient to support ATP hydrolysis and translocation in an ABC transporter, the histidine permease.* J Biol Chem, 1999. **274**(38): p. 26727-35.

109. Muir, M., L. Williams, and T. Ferenci, *Influence of transport energization on the growth yield of Escherichia coli.* J Bacteriol, 1985. **163**(3): p. 1237-42.

110. Mimmack, M.L., et al., *Energy coupling to periplasmic binding protein-dependent transport systems: stoichiometry of ATP hydrolysis during transport in vivo.* Proc Natl Acad Sci U S A, 1989. **86**(21): p. 8257-61.

111. Davidson, A.L. and H. Nikaido, *Overproduction, solubilization, and reconstitution of the maltose transport system from Escherichia coli.* J Biol Chem, 1990. **265**(8): p. 4254-60.

112. Pardee, A.B., *Membrane transport proteins. Proteins that appear to be parts of membrane transport systems are being isolated and characterized.* Science, 1968. **162**(854): p. 632-7.

113. Shuman, H.A., *Active transport of maltose in Escherichia coli K12. Role of the periplasmic maltose-binding protein and evidence for a substrate recognition site in the cytoplasmic membrane.* J Biol Chem, 1982. **257**(10): p. 5455-61.

114. Brass, J.M., W. Boos, and R. Hengge, *Reconstitution of maltose transport in malB mutants of Escherichia coli through calcium-induced disruptions of the outer membrane.* J Bacteriol, 1981. **146**(1): p. 10-7.

115. van der Heide, T. and B. Poolman, *ABC transporters: one, two or four extracytoplasmic substrate-binding sites?* Embo Rep, 2002. **3**(10): p. 938-43.

116. Quiocho, F.A., *Atomic basis of the exquisite specificity of phosphate and sulfate transport receptors.* Kidney Int, 1996. **49**(4): p. 943-6.

117. Higgins, C.F. and G.F. Ames, *Two periplasmic transport proteins which interact with a common membrane receptor show extensive homology: complete nucleotide sequences.* Proc Natl Acad Sci U S A, 1981. **78**(10): p. 6038-42.

118. Kempf, B., J. Gade, and E. Bremer, *Lipoprotein from the osmoregulated ABC transport system OpuA of Bacillus subtilis: purification of the glycine betaine binding protein and characterization of a functional lipidless mutant.* J Bacteriol, 1997. **179**(20): p. 6213-20.

119. Levdikov, V.M., et al., *The structure of the oligopeptide-binding protein, AppA, from Bacillus subtilis in complex with a nonapeptide.* J Mol Biol, 2005. **345**(4): p. 879-92.

120. Yang, X., et al., *Crystal structure of lipoprotein GNA1946 from Neisseria meningitidis.* J Struct Biol, 2009. **168**(3): p. 437-43.

121. Quiocho, F.A. and P.S. Ledvina, *Atomic structure and specificity of bacterial periplasmic receptors for active transport and chemotaxis: variation of common themes.* Mol Microbiol, 1996. **20**(1): p. 17-25.

122. Tame, J.R., et al., *The structural basis of sequence-independent peptide binding by OppA protein.* Science, 1994. **264**(5165): p. 1578-81.

123. Nickitenko, A.V., S. Trakhanov, and F.A. Quiocho, *2 A resolution structure of DppA, a periplasmic dipeptide transport/chemosensory receptor.* Biochemistry, 1995. **34**(51): p. 16585-95.

124. Heddle, J., et al., *Crystal structures of the liganded and unliganded nickel-binding protein NikA from Escherichia coli.* J Biol Chem, 2003. **278**(50): p. 50322-9.

125. Trakhanov, S., et al., *Ligand-free and -bound structures of the binding protein (LivJ) of the Escherichia coli ABC leucine/isoleucine/valine transport system: trajectory and dynamics of the interdomain rotation and ligand specificity.* Biochemistry, 2005. **44**(17): p. 6597-608.

126. Mao, B., et al., *Hinge-bending in L-arabinose-binding protein. The "Venus's-flytrap" model.* J Biol Chem, 1982. **257**(3): p. 1131-3.

127. Berntsson, R.P., et al., *A structural classification of substrate-binding proteins.* FEBS Lett, 2010.

128. Wilkinson, A.J. and K.H.G. Verschueren, *Crystal structures of periplasmic solute-binding proteins in ABC transport complexes illuminate their function*, in *ABC proteins: from bacteria to man.*, I.B. Holland, et al., Editors. 2003, Academic Press: London, United Kingdom. p. 187–207.

129. Quiocho, F.A., J.C. Spurlino, and L.E. Rodseth, *Extensive features of tight oligosaccharide binding revealed in high-resolution structures of the maltodextrin transport/chemosensory receptor.* Structure, 1997. **5**(8): p. 997-1015.

130. Sharff, A.J., et al., *Crystallographic evidence of a large ligand-induced hinge-twist motion between the two domains of the maltodextrin binding protein involved in active transport and chemotaxis.* Biochemistry, 1992. **31**(44): p. 10657-63.

131. Borths, E.L., et al., *The structure of Escherichia coli BtuF and binding to its cognate ATP binding cassette transporter.* Proc Natl Acad Sci U S A, 2002. **99**(26): p. 16642-7.

132. Sebulsky, M.T., et al., *The role of FhuD2 in iron(III)-hydroxamate transport in Staphylococcus aureus. Demonstration that FhuD2 binds iron(III)-hydroxamates but with minimal conformational change and implication of mutations on transport.* J Biol Chem, 2003. **278**(50): p. 49890-900.

133. Treptow, N.A. and H.A. Shuman, *Allele-specific malE mutations that restore interactions between maltose-binding protein and the inner-membrane components of the maltose transport system.* J Mol Biol, 1988. **202**(4): p. 809-22.

134. Szmelcman, S., N. Sassoon, and M. Hofnung, *Residues in the alpha helix 7 of the bacterial maltose binding protein which are important in interactions with the Mal FGK2 complex.* Protein Sci, 1997. **6**(3): p. 628-36.

135. Jones, P.M., M.L. O'Mara, and A.M. George, *ABC transporters: a riddle wrapped in a mystery inside an enigma.* Trends Biochem Sci, 2009. **34**(10): p. 520-31.

136. Janas, E., et al., *The ATP hydrolysis cycle of the nucleotide-binding domain of the mitochondrial ATP-binding cassette transporter Mdl1p.* J Biol Chem, 2003. **278**(29): p. 26862-9.

137. van der Does, C. and R. Tampe, *How do ABC transporters drive transport?* Biol Chem, 2004. **385**(10): p. 927-33.

138. Higgins, C.F. and K.J. Linton, *The ATP switch model for ABC transporters.* Nat Struct Mol Biol, 2004. **11**(10): p. 918-26.

139. Senior, A.E., M.K. al-Shawi, and I.L. Urbatsch, *The catalytic cycle of P-glycoprotein.* FEBS Lett, 1995. **377**(3): p. 285-9.

140. Sauna, Z.E., et al., *Catalytic cycle of ATP hydrolysis by P-glycoprotein: evidence for formation of the E.S reaction intermediate with ATP-gamma-S, a nonhydrolyzable analogue of ATP.* Biochemistry, 2007. **46**(48): p. 13787-99.

141. Westfahl, K.M., et al., *Functionally important ATP binding and hydrolysis sites in Escherichia coli MsbA.* Biochemistry, 2008. **47**(52): p. 13878-86.

142. Khare, D., et al., *Alternating access in maltose transporter mediated by rigid-body rotations.* Mol Cell, 2009. **33**(4): p. 528-36.

143. Higgins, C.F., *Multiple molecular mechanisms for multidrug resistance transporters.* Nature, 2007. **446**(7137): p. 749-57.

144. Davidson, A.L., *Mechanism of coupling of transport to hydrolysis in bacterial ATP-binding cassette transporters.* J Bacteriol, 2002. **184**(5): p. 1225-33.

145. Urbatsch, I.L., et al., *Both P-glycoprotein nucleotide-binding sites are catalytically active.* J Biol Chem, 1995. **270**(45): p. 26956-61.

146. Qu, Q., P.L. Russell, and F.J. Sharom, *Stoichiometry and affinity of nucleotide binding to P-glycoprotein during the catalytic cycle.* Biochemistry, 2003. **42**(4): p. 1170-7.

147. Buchaklian, A.H. and C.S. Klug, *Characterization of the Walker A motif of MsbA using site-directed spin labeling electron paramagnetic resonance spectroscopy.* Biochemistry, 2005. **44**(14): p. 5503-9.

148. Buchaklian, A.H. and C.S. Klug, *Characterization of the LSGGQ and H motifs from the Escherichia coli lipid A transporter MsbA.* Biochemistry, 2006. **45**(41): p. 12539-46.

149. Zolnerciks, J.K., C. Wooding, and K.J. Linton, *Evidence for a Sav1866-like architecture for the human multidrug transporter P-glycoprotein.* Faseb J, 2007. **21**(14): p. 3937-48.

150. Grote, M., et al., *A comparative electron paramagnetic resonance study of the nucleotide-binding domains' catalytic cycle in the assembled maltose ATP-binding cassette importer.* Biophys J, 2008. **95**(6): p. 2924-38.

151. Kennedy, K.A. and B. Traxler, *MalK forms a dimer independent of its assembly into the MalFGK2 ATP-binding cassette transporter of Escherichia coli.* J Biol Chem, 1999. **274**(10): p. 6259-64.

152. Hobson, A.C., R. Weatherwax, and G.F. Ames, *ATP-binding sites in the membrane components of histidine permease, a periplasmic transport system.* Proc Natl Acad Sci U S A, 1984. **81**(23): p. 7333-7.

153. Bishop, L., et al., *Reconstitution of a bacterial periplasmic permease in proteoliposomes and demonstration of ATP hydrolysis concomitant with transport.* Proc Natl Acad Sci U S A, 1989. **86**(18): p. 6953-7.

154. van Veen, H.W., et al., *The homodimeric ATP-binding cassette transporter LmrA mediates multidrug transport by an alternating two-site (two-cylinder engine) mechanism.* Embo J, 2000. **19**(11): p. 2503-14.

155. Sauna, Z.E. and S.V. Ambudkar, *Evidence for a requirement for ATP hydrolysis at two distinct steps during a single turnover of the catalytic cycle of human P-glycoprotein.* Proc Natl Acad Sci U S A, 2000. **97**(6): p. 2515-20.

156. Martin, C., et al., *Drug binding sites on P-glycoprotein are altered by ATP binding prior to nucleotide hydrolysis.* Biochemistry, 2000. **39**(39): p. 11901-6.

157. Liu, R. and F.J. Sharom, *Fluorescence studies on the nucleotide binding domains of the P-glycoprotein multidrug transporter.* Biochemistry, 1997. **36**(10): p. 2836-43.

158. Loo, T.W. and D.M. Clarke, *Drug-stimulated ATPase activity of human P-glycoprotein is blocked by disulfide cross-linking between the nucleotide-binding sites.* J Biol Chem, 2000. **275**(26): p. 19435-8.

159. Austermuhle, M.I., et al., *Maltose-binding protein is open in the catalytic transition state for ATP hydrolysis during maltose transport.* J Biol Chem, 2004. **279**(27): p. 28243-50.

160. Hall, J.A., et al., *Two modes of ligand binding in maltose-binding protein of Escherichia coli. Functional significance in active transport.* J Biol Chem, 1997. **272**(28): p. 17615-22.

161. Ambudkar, S.V., et al., *Relation between the turnover number for vinblastine transport and for vinblastine-stimulated ATP hydrolysis by human P-glycoprotein.* J Biol Chem, 1997. **272**(34): p. 21160-6.

162. Campbell, J.D., et al., *Extending the structure of an ABC transporter to atomic resolution: modeling and simulation studies of MsbA.* Biochemistry, 2003. **42**(13): p. 3666-73.

163. Tanizaki, S. and M. Feig, *Molecular dynamics simulations of large integral membrane proteins with an implicit membrane model.* J Phys Chem B, 2006. **110**(1): p. 548-56.

164. Haubertin, D.Y., et al., *Molecular dynamics simulations of E. coli MsbA transmembrane domain: formation of a semipore structure.* Biophys J, 2006. **91**(7): p. 2517-31.

165. Sonne, J., et al., *Simulation of the coupling between nucleotide binding and transmembrane domains in the ATP binding cassette transporter BtuCD.* Biophys J, 2007. **92**(8): p. 2727-34.

166. Ivetac, A. and M.S. Sansom, *Molecular dynamics simulations and membrane protein structure quality.* Eur Biophys J, 2008. **37**(4): p. 403-9.

167. Weng, J., et al., *Asymmetric conformational flexibility in the ATP-binding cassette transporter HI1470/1.* Biophys J, 2009. **96**(5): p. 1918-30.

168. Aittoniemi, J., et al., *Asymmetric switching in a homodimeric ABC transporter: a simulation study.* PLoS Comput Biol, 2010. **6**(4): p. e1000762.

169. Kandt, C. and D.P. Tieleman, *Holo-BtuF stabilizes the open conformation of the vitamin B12 ABC transporter BtuCD.* Proteins, 2010. **78**(3): p. 738-53.

170. Weng, J.W., K.N. Fan, and W.N. Wang, *The conformational transition pathway of ATP-binding cassette transporter MsbA revealed by atomistic simulations.* J Biol Chem, 2010. **285**(5): p. 3053-63.

171. Campbell, J.D., et al., *Nucleotide-dependent conformational changes in HisP: molecular dynamics simulations of an ABC transporter nucleotide-binding domain.* Biophys J, 2004. **87**(6): p. 3703-15.

172. Campbell, J.D. and M.S. Sansom, *Nucleotide binding to the homodimeric MJ0796 protein: a computational study of a prokaryotic ABC transporter NBD dimer.* FEBS Lett, 2005. **579**(19): p. 4193-9.

173. Jones, P.M. and A.M. George, *Mechanism of ABC transporters: a molecular dynamics simulation of a well characterized nucleotide-binding subunit.* Proc Natl Acad Sci U S A, 2002. **99**(20): p. 12639-44.

174. Kerr, I.D., *Structure and association of ATP-binding cassette transporter nucleotide-binding domains.* Biochim Biophys Acta, 2002. **1561**(1): p. 47-64.

175. Seigneuret, M. and A. Garnier-Suillerot, *A structural model for the open conformation of the mdr1 P-glycoprotein based on the MsbA crystal structure.* J Biol Chem, 2003. **278**(32): p. 30115-24.

176. DeGorter, M.K., et al., *Molecular modeling of the human multidrug resistance protein 1 (MRP1/ABCC1).* Biochem Biophys Res Commun, 2008. **365**(1): p. 29-34.

177. Hazai, E. and Z. Bikadi, *Homology modeling of breast cancer resistance protein (ABCG2).* J Struct Biol, 2008. **162**(1): p. 63-74.

178. Mornon, J.P., P. Lehn, and I. Callebaut, *Molecular models of the open and closed states of the whole human CFTR protein.* Cell Mol Life Sci, 2009. **66**(21): p. 3469-86.

179. Becker, J.P., et al., *Molecular models of human P-glycoprotein in two different catalytic states.* BMC Struct Biol, 2009. **9**: p. 3.

180. Ames, G.F., et al., *Traffic ATPases: a superfamily of transport proteins operating from Escherichia coli to humans.* Adv Enzymol Relat Areas Mol Biol, 1992. **65**: p. 1-47.

181. Ferenci, T., *The recognition of maltodextrins by Escherichia coli.* Eur J Biochem, 1980. **108**(2): p. 631-6.

182. Davidson, A.L. and H. Nikaido, *Purification and characterization of the membrane-associated components of the maltose transport system from Escherichia coli.* J Biol Chem, 1991. **266**(14): p. 8946-51.

183. Oldham, M.L., A.L. Davidson, and J. Chen, *Structural insights into ABC transporter mechanism.* Curr Opin Struct Biol, 2008. **18**(6): p. 726-33.

184. Shapiro, A.B. and V. Ling, *Reconstitution of drug transport by purified P-glycoprotein.* J Biol Chem, 1995. **270**(27): p. 16167-75.

185. Serohijos, A.W., et al., *Phenylalanine-508 mediates a cytoplasmic-membrane domain contact in the CFTR 3D structure crucial to assembly and channel function.* Proc Natl Acad Sci U S A, 2008. **105**(9): p. 3256-61.

186. Lawson, J., M.L. O'Mara, and I.D. Kerr, *Structure-based interpretation of the mutagenesis database for the nucleotide binding domains of P-glycoprotein.* Biochim Biophys Acta-Biomembranes, 2008. **1778**(2): p. 376-391.

187. Procko, E., et al., *The mechanism of ABC transporters: general lessons from structural and functional studies of an antigenic peptide transporter.* Faseb J, 2009. **23**(5): p. 1287-302.

188. Bult, C.J., et al., *Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii.* Science, 1996. **273**(5278): p. 1058-73.

189. Moody, J.E. and P.J. Thomas, *Nucleotide binding domain interactions during the mechanochemical reaction cycle of ATP-binding cassette transporters.* J Bioenerg Biomembr, 2005. **37**(6): p. 475-9.

190. Leach, A.R., *Molecular Modelling: Principles and Applications*. 1996, England: LONGMAN.

191. van Gunsteren, W.F., et al., *Biomolecular modeling: Goals, problems, perspectives.* Angew Chem Int Ed Engl, 2006. **45**(25): p. 4064-92.

192. van Gunsteren, W.F. and H.J.C. Berendsen, *Computer simulation of molecular dynamics: methodology, applications, and perspectives in chemistry.* Angew Chem Int, 1990. **29**: p. 992–1023.

193. Frenkel, D. and S. Berend, *Understanding Molecular Simulation: From Algorithms to Applications*. second edition ed. Computational Science Series: Volume 1. 2002: Academic Press.

194. Berendsen, H.J.C., *Simulating the Physical World: Hierarchical Modeling from Quantum Mechanics to Fluid Dynamics*. 2007: Cambridge University Press.

195. Kaminski, G.A., et al., *Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides.* J Phys Chem B, 2001. **105**(28): p. 6474-6487.

196. van Gunsteren, W.F., et al., *Biomolecular simulation: The GROMOS96 manual and user guide.* 1996, Zurich, Groninger: BIOMOS b.v.

197. Oostenbrink, C., et al., *A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6.* J Comput Chem, 2004. **25**(13): p. 1656-76.

198. Marrink, S.J., A.H. de Vries, and A.E. Mark, *Coarse grained model for semiquantitative lipid simulations.* J Phys Chem B, 2004. **108**(2): p. 750-760.

199. Scott, W.R.P., et al., *The GROMOS biomolecular simulation program package.* J Phys Chem, 1999. **103**: p. 3596-3607.

200. Oostenbrink, C., et al., *Validation of the 53A6 GROMOS force field.* Eur Biophys J, 2005. **34**(4): p. 273-84.

201. Schreiber, H. and O. Steinhauser, *Cutoff size does strongly influence molecular dynamics results on solvated polypeptides.* Biochemistry, 1992. **31**(25): p. 5856-60.

202. Darden, T., et al., *New tricks for modelers from the crystallography toolkit: the particle mesh Ewald algorithm and its use in nucleic acid simulations.* Structure, 1999. **7**(3): p. R55-60.

203. Darden, T., D. York, and L. Pedersen, *Particle mesh Ewald: an N.log(N) method for ewald sums in large systems.* J Chem Phys, 1993. **98**: p. 10089-10092.

204. Tironi, I.G., et al., *A generalized reaction field method for molecular-dynamics simulations.* J Chem Phys, 1995. **102**: p. 15451-5459.

205. Essmann, U., L. Perera, and M.L. Berkowitz, *A smooth particle mesh Ewald method.* J Chem Phys, 1995. **103**: p. 8577-8593.

206. Bergdorf, M., C. Peter, and P.H. Hunenberger, *Influence of cut-off truncation and artificial periodicity of electrostatic interactions in molecular simulations of solvated ions: A continuum electrostatics study.* J Chem Phys, 2003. **119**(17): p. 9129-9144.

207. Adcock, S.A. and J.A. McCammon, *Molecular dynamics: survey of methods for simulating the activity of proteins.* Chem Rev, 2006. **106**(5): p. 1589-615.

208. Bashford, D. and D.A. Case, *Generalized Born Models of Macromolecular Solvation Effects.* Annu Rev Phys Chem, 2000. **51**: p. 129- 152.

209. Hunenberger, P.H., *Molecular dynamics simulations using empirical force fields: Principles and applications to selected systems of chemical and biochemical interest*. 1997: ETH Zürich.

210. Verlet, L., *Computer Experiments on Classical Fluids .I. Thermodynamical Properties of Lennard-Jones Molecules.* Phys Rev, 1967. **159**(1): p. 98-&.

211. Hockney, R.W., S.P. Goel, and J.W. Eastwood, *Quiet High-Resolution Computer Models of a Plasma.* J Comput Phys, 1974. **14**(2): p. 148-158.

212. Press, W., et al. 1986, Cambridge, England: Cambridge University Press.

213. Lindahl, E., B. Hess, and D. van der Spoel, *GROMACS 3.0: a package for molecular simulation and trajectory analysis.* J Mol Model, 2001. **7**: p. 306-317.

214. Van Der Spoel, D., et al., *GROMACS: fast, flexible, and free.* J Comput Chem, 2005. **26**(16): p. 1701-18.

215. Ryckaert, J.P., G. Ciccotti, and H.J.C. Berendsen, *Numerical integration of the cartesian equations of motion of a system with constraints; molecular dynamics of n-alkanes.* J Comp Phys, 1977. **23**: p. 327–341.

216. Hess, B., et al., *LINCS: A Linear Constraint Solver for molecular simulations.* J Comp Chem, 1997. **18**: p. 1463-1472.

217. Miyamoto, S. and P.A. Kollman, *SETTLE: An Analytical Version of the SHAKE and RATTLE Algorithms for Rigid Water Models.* J Comp Chem, 1992. **13**: p. 952-962.

218. Metropolis, N., et al., *Equation of state calculations by fast computing machines.* J Chem Phys 1953. **21**(6): p. 1087-1092.

219. Andersen, O.S. and R.E. Koeppe, 2nd, *Bilayer thickness and membrane protein function: an energetic perspective.* Annu Rev Biophys Biomol Struct, 2007. **36**: p. 107-30.

220. McElhaney, R.N., *The influence of membrane lipid composition and physical properties of membrane structure and function in Acholeplasma laidlawii.* Crit Rev Microbiol, 1989. **17**(1): p. 1-32.

221. Ash, W.L., et al., *Computer simulations of membrane proteins.* Biochim Biophys Acta, 2004. **1666**(1-2): p. 158-89.

222. Roux, B., B. Prod'hom, and M. Karplus, *Ion transport in the gramicidin channel: molecular dynamics study of single and double occupancy.* Biophys J, 1995. **68**(3): p. 876-92.

223. Zhong, Q., et al., *Molecular dynamics simulation of a synthetic ion channel.* Biophys J, 1998. **74**(1): p. 3-10.

224. Im, W., M. Feig, and C.L. Brooks, 3rd, *An implicit membrane generalized born theory for the study of structure, stability, and interactions of membrane proteins.* Biophys J, 2003. **85**(5): p. 2900-18.

225. Kessel, A., T. Haliloglu, and N. Ben-Tal, *Interactions of the M2delta segment of the acetylcholine receptor with lipid bilayers: a continuum-solvent model study.* Biophys J, 2003. **85**(6): p. 3687-95.

226. Scott, H.L., *Modeling the lipid component of membranes.* Curr Opin Struct Biol, 2002. **12**(4): p. 495-502.

227. Fogolari, F., et al., *Biomolecular electrostatics with the linearized Poisson-Boltzmann equation.* Biophys J, 1999. **76**(1 Pt 1): p. 1-16.

228. Fogolari, F., A. Brigo, and H. Molinari, *The Poisson-Boltzmann equation for biomolecular electrostatics: a tool for structural biology.* J Mol Recognit, 2002. **15**(6): p. 377-92.

229. Roux, B. and T. Simonson, *Implicit solvent models.* Biophys Chem, 1999. **78**(1-2): p. 1-20.

230. Honig, B. and A. Nicholls, *Classical electrostatics in biology and chemistry.* Science, 1995. **268**(5214): p. 1144-9.

231. Fogolari, F., et al., *Molecular mechanics and dynamics of biomolecules using a solvent continuum model.* J Comput Chem, 2001. **22**(15): p. 1830-1842.

232. Warwicker, J. and H.C. Watson, *Calculation of the Electric-Potential in the Active-Site Cleft Due to Alpha-Helix Dipoles.* J Mol Biol, 1982. **157**(4): p. 671-679.

233. Martel, P., A. Baptista, and S.B. Petersen, *Protein electrostatics*, in *Biotechnology. Annual Review* 1996. p. 315–372.

234. Bashford, D. and K. Gerwert, *Electrostatic Calculations of the pKa Values of Ionizable Groups in Bacteriorhodopsin.* J Mol Biol, 1992. **224**: p. 473-486.

235. Yang, A.S., et al., *On the calculation of pKas in proteins.* Proteins, 1993. **15**(3): p. 252-65.

236. Gilson, M.K., *Multiple-site titration and molecular modeling: two rapid methods for computing energies and forces for ionizable groups in proteins.* Proteins, 1993. **15**(3): p. 266-82.

237. Antosiewicz, J., J.A. McCammon, and M.K. Gilson, *Prediction of pH-dependent properties of proteins.* J Mol Biol, 1994. **238**(3): p. 415-36.

238. Schaefer, M., M. Sommer, and M. Karplus, *pH-dependence of protein stability: Absolute electrostatic free energy differences between conformations.* J Phys Chem B, 1997. **101**(9): p. 1663-1683.

239. Baptista, A.M. and C.M. Soares, *Some theoretical and computational aspects of the inclusion of proton isomerism in the protonation equilibrium of proteins.* J Phys Chem, 2001. **105**: p. 293-309.

240. Bashford, D. and M. Karplus, *pKa's of ionizable groups in proteins: atomic detail from a continuum electrostatic model.* Biochemistry, 1990. **29**(44): p. 10219-25.

241. Beroza, P., et al., *Protonation of interacting residues in a protein by a Monte Carlo method: application to lysozyme and the photosynthetic reaction center of Rhodobacter sphaeroides.* Proc Natl Acad Sci U S A, 1991. **88**(13): p. 5804-8.

242. Klapper, I., et al., *Focusing of electric fields in the active site of Cu-Zn superoxide dismutase: effects of ionic strength and amino-acid modification.* Proteins, 1986. **1**(1): p. 47-59.

243. Teixeira, V.H., et al., *On the use of different dielectric constants for computing individual and pairwise terms in poisson-boltzmann studies of protein ionization equilibrium.* J Phys Chem B, 2005. **109**(30): p. 14691-706.

244. Albers, S.V., et al., *Insights into ABC transport in archaea.* J Bioenerg Biomembr, 2004. **36**(1): p. 5-15.

245. Gottesman, M.M. and S.V. Ambudkar, *Overview: ABC transporters and human disease.* J Bioenerg Biomembr, 2001. **33**(6): p. 453-8.

246. Deeley, R.G., C. Westlake, and S.P. Cole, *Transmembrane transport of endo- and xenobiotics by mammalian ATP-binding cassette multidrug resistance proteins.* Physiol Rev, 2006. **86**(3): p. 849-99.

247. Higgins, C.F., *ABC transporters: physiology, structure and mechanism--an overview.* Res Microbiol, 2001. **152**(3-4): p. 205-10.

248. Linton, K.J., *Structure and function of ABC transporters.* Physiology (Bethesda), 2007. **22**: p. 122-30.

249. Schmitt, L. and R. Tampe, *Structure and mechanism of ABC transporters.* Curr Opin Struct Biol, 2002. **12**(6): p. 754-60.

250. Gerber, S., et al., *Structural basis of trans-inhibition in a molybdate/tungstate ABC transporter.* Science, 2008. **321**(5886): p. 246-50.

251. Fetsch, E.E. and A.L. Davidson, *Vanadate-catalyzed photocleavage of the signature motif of an ATP-binding cassette (ABC) transporter.* Proc Natl Acad Sci U S A, 2002. **99**(15): p. 9685-90.

252. Delano, W., *The Pymol Molecular Graphics System. version 0.98*. 2003, San Carlos, CA, USA: Delano Scientific LLC.

253. Frauenfelder, H., S.G. Sligar, and P.G. Wolynes, *The energy landscapes and motions of proteins.* Science, 1991. **254**: p. 1598-1603.

254. Oliveira, A.S., et al., *Reorganization and conformational changes in the reduction of tetraheme cytochromes.* Biophys J, 2005. **89**(6): p. 3919-30.

255. Teixeira, V.H., C.M. Soares, and A.M. Baptista, *Studies of the reduction and protonation behavior of tetraheme cytochromes using atomic detail.* J Biol Inorg Chem, 2002. **7**: p. 200-216.

256. Bashford, D., *An Object-Oriented Programming Suite for Electrostatic Effects in Biological Molecules*, in *Scientific Computing in Object-Oriented Parallel Environments*, Y. Ishikawa, et al., Editors. 1997, ISCOPE97, Springer: Berlin. p. 233-240.

257. Berendsen, H., D. Spoel, and R. Drunen, *GROMACS 3.1.4.* Comp Phys Comm, 1995. **91**: p. 43-56.

258. Hermans, J., et al., *A Consistent Empirical Potential for Water-Protein Interactions.* Biopolymers, 1984. **23**(8): p. 1513-1518.

259. Eisenhaber, F. and P. Argos, *Improved Strategy in Analytic Surface Calculation for Molecular-Systems - Handling of Singularities and Computational Efficiency.* J Comput Chem, 1993. **14**: p. 1272-1280.

260. Eisenhaber, F., et al., *The Double Cubic Lattice Method - Efficient Approaches to Numerical-Integration of Surface-Area and Volume and to Dot Surface Contouring of Molecular Assemblies.* J Comput Chem, 1995. **16**: p. 273-284.

261. Berendsen, H., et al., *Molecular dynamics with coupling to an external bath.* J Chem Phys, 1984. **81**: p. 3684-3690.

262. Kabsch, W. and C. Sander, *Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features.* Biopolymers, 1983. **22**(12): p. 2577-637.

263. Ciccotti, G., G. Jacucci, and I.R. McDonald, *"Thought-Experiments" by Molecular Dynamics.* J Stat Phys, 1979. **21**: p. 1-21.

264. Wang, C., et al., *Dynamics of ATP-binding cassette contribute to allosteric control, nucleotide binding and energy transduction in ABC transporters.* J Mol Biol, 2004. **342**(2): p. 525-37.

265. Waters, M.L., *Aromatic interactions in model systems.* Curr Opin Chem Biol, 2002. **6**(6): p. 736-41.

266. Allen, M.P. and D.J. Tildesley, *Computer simulation of liquids*. 1987, Oxford, UK: Clarendon Press.

267. Mourez, M., et al., *In vitro interaction between components of the inner membrane complex of the maltose ABC transporter of Escherichia coli: modulation by ATP.* Mol Microbiol, 1998. **30**(2): p. 353-63.

268. Oancea, G., et al., *Structural arrangement of the transmission interface in the antigen ABC transport complex TAP.* Proc Natl Acad Sci U S A, 2009. **106**(14): p. 5551-6.

269. He, L., et al., *Multiple membrane-cytoplasmic domain contacts in the cystic fibrosis transmembrane conductance regulator (CFTR) mediate regulation of channel gating.* J Biol Chem, 2008. **283**(39): p. 26383-90.

270. Dean, M., A. Rzhetsky, and R. Allikmets, *The human ATP-binding cassette (ABC) transporter superfamily.* Genome Res, 2001. **11**(7): p. 1156-66.

271. Linton, K.J. and C.F. Higgins, *The Escherichia coli ATP-binding cassette (ABC) proteins.* Mol Microbiol, 1998. **28**(1): p. 5-13.

272. Davidson, A.L. and P.C. Maloney, *ABC transporters: how small machines do a big job.* Trends Microbiol, 2007. **15**(10): p. 448-55.

273. Saurin, W., M. Hofnung, and E. Dassa, *Getting in or out: early segregation between importers and exporters in the evolution of ATP-binding cassette (ABC) transporters.* J Mol Evol, 1999. **48**(1): p. 22-41.

274. Kandt, C., W.L. Ash, and D.P. Tieleman, *Setting up and running molecular dynamics simulations of membrane proteins.* Methods, 2007. **41**(4): p. 475-88.

275. Chiu, S.W., et al., *Incorporation of surface tension into molecular dynamics simulation of an interface: a fluid phase lipid bilayer membrane.* Biophys J, 1995. **69**(4): p. 1230-45.

276. Koyama, T.M., et al., *Characterizing the Gel to Liquid Crystal Transition in Lipid-Bilayer Model Systems.* Chem Educator, 1999. **4**(1): p. 12-15.

277. Chandrasekhar, I., et al., *Molecular dynamics simulation of lipid bilayers with GROMOS96: Application of surface tension.* Mol Simulat, 2005. **31**(8): p. 543-548.

278. Barker, J.A. and R.O. Watts, *Monte-Carlo Studies of Dielectric Properties of Water-Like Models.* Mol Phys, 1973. **26**(3): p. 789-792.

279. Frezard, F., et al., *P-glycoprotein preferentially effluxes anthracyclines containing free basic versus charged amine.* Eur J Biochem, 2001. **268**(6): p. 1561-7.

280. Siarheyeva, A., J.J. Lopez, and C. Glaubitz, *Localization of multidrug transporter substrates within model membranes.* Biochemistry, 2006. **45**(19): p. 6203-11.

281. Frisch, M.J., et al., *Gaussian 03, Revision C.02*. 2004, Wallingford, CT: Gaussian, Inc.

282. Bayly, C.I., et al., *A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges - the Resp Model.* J Phys Chem, 1993. **97**: p. 10269-10280.

283. Morris, G.M., et al., *AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility.* J Comput Chem, 2009. **30**(16): p. 2785-91.

284. Zaitseva, J., et al., *A molecular understanding of the catalytic cycle of the nucleotide-binding domain of the ABC transporter HlyB.* Biochem Soc Trans, 2005. **33**(Pt 5): p. 990-5.

285. Smith, C.A. and I. Rayment, *X-ray structure of the magnesium(II).ADP.vanadate complex of the Dictyostelium discoideum myosin motor domain to 1.9 A resolution.* Biochemistry, 1996. **35**(17): p. 5404-17.

286. Dawson, R.J., K. Hollenstein, and K.P. Locher, *Uptake or extrusion: crystal structures of full ABC transporters suggest a common mechanism.* Mol Microbiol, 2007. **65**(2): p. 250-7.

287. Kuhnau, S., et al., *The activities of the Escherichia coli MalK protein in maltose transport, regulation, and inducer exclusion can be separated by mutations.* J Bacteriol, 1991. **173**(7): p. 2180-6.

288. Nelson, S.O. and P.W. Postma, *Interactions in vivo between IIIGlc of the phosphoenolpyruvate:sugar phosphotransferase system and the glycerol and maltose uptake systems of Salmonella typhimurium.* Eur J Biochem, 1984. **139**(1): p. 29-34.

289. Dean, D.A., et al., *Regulation of the maltose transport system of Escherichia coli by the glucose-specific enzyme III of the phosphoenolpyruvate-sugar phosphotransferase system. Characterization of inducer exclusion-resistant mutants and reconstitution of inducer exclusion in proteoliposomes.* J Biol Chem, 1990. **265**(34): p. 21005-10.

290. Stein, A., et al., *Functional characterization of the maltose ATP-binding-cassette transporter of Salmonella typhimurium by means of monoclonal antibodies directed against the Ma1K subunit.* Eur J Biochem, 2002. **269**(16): p. 4074-4085.

291. Sali, A. and T.L. Blundell, *Comparative Protein Modeling by Satisfaction of Spatial Restraints.* J Mol Biol, 1993. **234**(3): p. 779-815.

292. Lins, R.D. and P.H. Hunenberger, *A new GROMOS force field for hexopyranose-based carbohydrates.* J Comput Chem, 2005. **26**(13): p. 1400-12.

293. Smith, P.E. and W.F. Vangunsteren, *Consistent Dielectric-Properties of the Simple Point-Charge and Extended Simple Point-Charge Water Models at 277 and 300 K.* J Chem Phys, 1994. **100**(4): p. 3169-3174.

294. Kollman, P., *Free-Energy Calculations - Applications to Chemical and Biochemical Phenomena.* Chem Rev, 1993. **93**(7): p. 2395-2417.

295. Hor, L.I. and H.A. Shuman, *Genetic analysis of periplasmic binding protein dependent transport in Escherichia coli. Each lobe of maltose-binding protein interacts with a different subunit of the MalFGK2 membrane transport complex.* J Mol Biol, 1993. **233**(4): p. 659-70.

296. Daus, M.L., et al., *Maltose binding protein (MalE) interacts with periplasmic loops P2 and P1 respectively of the MalFG subunits of the maltose ATP binding cassette transporter (MalFGK(2)) from Escherichia coli/Salmonella during the transport cycle.* Mol Microbiol, 2007. **66**(5): p. 1107-22.

297. Hyde, S.C., et al., *Structural model of ATP-binding proteins associated with cystic fibrosis, multidrug resistance and bacterial transport.* Nature, 1990. **346**(6282): p. 362-5.

298. Mourez, M., N. Hofnung, and E. Dassa, *Subunit interactions in ABC transporters: A conserved sequence in hydrophobic membrane proteins of periplasmic permeases defines an important site of interaction with the ATPase subunits.* Embo J, 1997. **16**(11): p. 3066-3077.

299. Peral, F. and E. Gallego, *The self-organization of adenosine 5'-triphosphate and adenosine 5'-diphosphate in aqueous solution as determined from ultraviolet hypochromic effects.* Biophys Chem, 2000. **85**(1): p. 79-92.

300. Sigel, H., *Isomeric equilibria in complexes of adenosine 5'-triphosphate with divalent metal ions. Solution structures of M(ATP)2- complexes.* Eur J Biochem, 1987. **165**(1): p. 65-72.

301. Sigel, H., et al., *Comparison of the stabilities of monomeric metal ion complexes formed with adenosine 5'-triphosphate (ATP)and pyrimidine-nucleoside 5´-triphosphate (CTP, UTP, TTP) and evaluation of the isomeric equilibria in the complexes of ATP and CTP.* Inorg Chem, 1987. **26**(13): p. 2149-2157.

302. Richter, Y. and B. Fischer, *Characterization and elucidation of coordination requirements of adenine nucleotides complexes with Fe(II) ions.* Nucleosides Nucleotides Nucleic Acids, 2003. **22**(9): p. 1757-80.

303. Goldberg, R.N., N. Kishore, and R.M. Lennen, *Thermodynamic quantities for the ionization reactions of buffers.* J Phys Chem Ref Data, 2002. **31**(2): p. 231-370.

304. Meagher, K.L., L.T. Redman, and H.A. Carlson, *Development of polyphosphate parameters for use with the AMBER force field.* J Comput Chem, 2003. **24**(9): p. 1016-25.

305. Hwang, M., et al., *Conformational Analysis of three pyrophosphate model species: Diphosphate, Methyl Diphosphate and Triphosphate.* J Comput Chem, 1999. **Vol.20**(16): p. 1702-1715.

306. Frisch, M.J., et al., *Gaussian 98, Revision A.7*. 1998, Gaussian, Inc.: Pittsburgh PA.

307. Berendsen, H.J.C., D. Vanderspoel, and R. Vandrunen, *Gromacs - a Message-Passing Parallel Molecular-Dynamics Implementation.* Comput Phys Commun, 1995. **91**(1-3): p. 43-56.

308. Nagle, J.F. and S. Tristram-Nagle, *Lipid bilayer structure.* Curr Opin Struc Biol, 2000. **10**(4): p. 474-480.

309. Petrache, H.I., S.W. Dodd, and M.F. Brown, *Area per lipid and acyl length distributions in fluid phosphatidylcholines determined by H-2 NMR spectroscopy.* Biophys J, 2000. **79**(6): p. 3172-3192.

310. Wohlert, J. and O. Edholm, *Dynamics in atomistic simulations of phospholipid membranes: Nuclear magnetic resonance relaxation rates and lateral diffusion.* J Chem Phys, 2006. **125**(20): p. -.

311. Lipscomb, L.A., et al., *Water ring structure at DNA interfaces: hydration and dynamics of DNA-anthracycline complexes.* Biochemistry, 1994. **33**(12): p. 3649-59.